

SEARCH REQUEST FORM

Requestor's

Name: _____

Serial

Number: _____

Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 02-14-02
 Searcher: Beverly E. 4994
 Terminal time: 40
 Elapsed time: _____
 CPU time: _____
 Total time: 57
 Number of Searches: _____
 Number of Databases: 3

Search Site

_____ STIC
 _____ CM-1
 _____ Pre-S

Type of Search

_____ N.A. Sequence
 _____ A.A. Sequence
 _____ Structure
 _____ Bibliographic

Vendors

_____ IG
 _____ ☒ STN
 _____ ☒ Dialog
 _____ APS
 _____ Geninfo
 _____ SDC
 _____ DARC/Questel
 _____ ☒ Other CGN

THIS PAGE BLANK (USPTO)

Devi, S.
09/360685

09/360685

FILE 'CAPLUS' ENTERED AT 10:21:36 ON 14 FEB 2002)

L1 6299 SEA FILE=CAPLUS ABB=ON PLU=ON PYLORI
L3 4 SEA FILE=CAPLUS ABB=ON PLU=ON L1 AND (CAI OR (CYTOTOXIN
OR CYTO TOXIN) (W) (ASSOC? OR ASS#) (W) (IMMUNODOMIN? OR
IMMUNO DOMIN?))

-key terms

L3 ANSWER 1 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 2000:687957 CAPLUS

DOCUMENT NUMBER: 133:247257

TITLE: Method and antisense conjugate for treating
Helicobacter pylori infection

INVENTOR(S): Iversen, Patrick L.; Brand, Randall; Weller,
Dwight D.; Summerton, James E.

PATENT ASSIGNEE(S): AVI BioPharma, Inc., USA; The Board of Regents
of the University of Nebraska

SOURCE: U.S., 21 pp., Cont.-in-part of U.S. 6,030,941.
CODEN: USXXAM

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 6124271	A	20000926	US 1998-12198	19980123
US 6030941	A	20000229	US 1997-848844	19970430
PRIORITY APPLN. INFO.:			US 1997-36366	P 19970124
			US 1997-848844	A2 19970430
			US 1996-16347	P 19960501
			US 1996-28609	P 19961023

AB A method and conjugate for treating H. pylori infection in
a subject are disclosed. The conjugate is composed of (a) a
nuclease-resistant antisense oligomer effective to inhibit H.
pylori infection in the subject by base-specific
Watson-Crick binding to an H. pylori mRNA transcript, and
(b) a transport moiety conjugated to the oligomer. The transport
moiety is effective to facilitate uptake of the conjugate from the
environment of the stomach into the cytoplasm of H. pylori
cells by active transport or by pH-gradient transport across of the
cell membrane of H. pylori. The conjugate is administered
by oral route, preferably in a swellable polymer bolus designed to
release the conjugate in sustained release.

REFERENCE COUNT: 8 THERE ARE 8 CITED REFERENCES AVAILABLE FOR
THIS RECORD. ALL CITATIONS AVAILABLE IN
THE RE FORMAT

L3 ANSWER 2 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1998:527235 CAPLUS

DOCUMENT NUMBER: 129:170512

TITLE: Antisense oligonucleotides conjugates taken up
Helicobacter pylori for treatment of
infection

INVENTOR(S): Iverson, Patrick L.; Bragg, Randall; Weller,
Dwight D.; Summerton, James E.

PATENT ASSIGNEE(S): Antivirals, Inc., USA; University of Nebraska
Board of Regents

SOURCE: PCT Int. Appl., 36 pp.
CODEN: PIXXD2

Searcher : Shears 308-4994

THIS PAGE BLANK (USPTO)

09/360685

DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9832467	A2	19980730	WO 1998-US1393	19980123
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
US 6030941	A	20000229	US 1997-848844	19970430
AU 9859305	A1	19980818	AU 1998-59305	19980123
EP 973886	A2	20000126	EP 1998-902712	19980123
R: AT, BE, CH, DE, ES, FR, GB, IT, LI, LU, NL				
JP 2001509167	T2	20010710	JP 1998-532174	19980123
PRIORITY APPLN. INFO.: US 1997-36366 P 19970124 US 1997-848844 A 19970430 US 1996-16347 P 19960501 US 1996-28609 P 19961023 WO 1998-US1393 W 19980123				

AB Antisense oligonucleotides inhibiting the expression of *Helicobacter pylori* genes that are conjugated with transport moieties that improve the efficiency of uptake by *H. pylori* are described for use in treatment of stomach infection. The transport moiety facilitates uptake from the environment of the stomach into the cytoplasm of *H. pylori* cells by active transport or by pH-gradient transport across of the cell membrane. The preferred transport moiety is a sugar such as D-galactose or L-arabinose that is taken by *Helicobacter* by active transport but that are not absorbed from the gut.

L3 ANSWER 3 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1996:747006 CAPLUS
 DOCUMENT NUMBER: 126:55630
 TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors
 AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello
 CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy
 SOURCE: Proc. Natl. Acad. Sci. U. S. A. (1996), 93(25), 14648-14653
 CODEN: PNASA6; ISSN: 0027-8424
 PUBLISHER: National Academy of Sciences
 DOCUMENT TYPE: Journal
 LANGUAGE: English
 AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with

Handwritten note:
 CagA gene...
 PNAS 93:14648-14653, 1996
 (Borodovsky) Aug.

THIS PAGE BLANK (USPTO)

09/360685

severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains *cagA* (*cag*) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, *cag* is split into a right segment (*cagI*) and a left segment (*cagII*) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, *cagI* and *cagII* are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the *cagI* region and the extreme 3' end of the *cagII* region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (*cagE*), which is similar to the toxin-secretion gene of *Bordetella pertussis*, *ptlC*, and the transport systems required for plasmid transfer, including the *virB4* gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the *cagI* genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the *cag* region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L3 ANSWER 4 OF 4 CAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1994:2256 CAPLUS

DOCUMENT NUMBER: 120:2256

TITLE: *Helicobacter pylori* proteins useful for vaccines and diagnostics

INVENTOR(S): Covacci, Antonello; Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino

PATENT ASSIGNEE(S): Biocine Sclavo S.p.A., Italy

SOURCE: PCT Int. Appl., 85 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9318150	A1	19930916	WO 1993-EP472	19930302
W:	AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US			
RW:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG			
AU 9336300	A1	19931005	AU 1993-36300	19930302
EP 643770	A1	19950322	EP 1993-905285	19930302
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE			
JP 07504565	T2	19950525	JP 1993-515309	19930302
EP 967279	A1	19991229	EP 1999-202698	19930302
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE			
JP 2000333686	A2	20001205	JP 2000-126696	19930302
JP 2000350591	A2	20001219	JP 2000-126695	19930302
US 6077706	A	20000620	US 1995-470260	19950606
US 6130059	A	20001010	US 1995-466662	19950606
PRIORITY APPLN. INFO.:			IT 1992-FI52	A 19920302
			WO 1993-EP158	A 19930125
			EP 1993-905285	A3 19930302

Searcher : Shears 308-4994

10/20/2011
10/20/2011
10/20/2011

10/20/2011
10/20/2011

THIS PAGE BLANK (USPTO)

09/360685

JP 1993-515309 A 19930302
WO 1993-EP472 A 19930302
US 1994-256848 B3 19941021

AB The H. **pylori** genes for cytotoxin, **CAI** (**cytotoxin-assocd. immunodominant**) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. **pylori**. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic H. **pylori** strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. **pylori** strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. **pylori** and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

(FILE 'MEDLINE, BIOSIS, EMBASE, WPIDS, JICST-EPLUS, JAPIO, TOXLIT, TOXCENTER, PHIC, PHIN' ENTERED AT 10:25:05 ON 14 FEB 2002)

L4
L5

8 S L3

1 DUP REM L4 (1 DUPLICATE REMOVED)

L5 ANSWER 1 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:239751 BIOSIS
DOCUMENT NUMBER: PREV200100239751
TITLE: Helicobacter **pylori** proteins useful for vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Vc.Provenzano, 8, 53100, Siena Italy
PATENT INFORMATION: US 6130059 October 10, 2000
SOURCE: Official Gazette of the United States Patent and Trademark Office Patents, (Oct. 10, 2000) Vol. 1239, No. 2, pp. No Pagination. e-file.
ISSN: 0098-1133.
DOCUMENT TYPE: Patent
LANGUAGE: English

AB Helicobacter **pylori** is known to cause or be a cofactor in type B gastritis, peptic ulcers, and gastric tumors. In both developed and developing countries, a high percentage of people are infected with this bacterium. The present invention relates generally to certain H. **pylori** proteins, to the genes which express these proteins, and to the use of these proteins for diagnostic and vaccine applications. Specifically, molecular cloning, nucleotide, and amino acid sequences for the H. **pylori** cytotoxin (CT), the "**Cytotoxin Associated Immunodominant**" (**CAI**) antigen, and the heat shock protein (hsp60). are described herein.

L5 ANSWER 2 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:181462 BIOSIS
DOCUMENT NUMBER: PREV200100181462
TITLE: Helicobacter **pylori** proteins useful for

THIS PAGE BLANK (USPTO)

09/360685

 vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford,
 John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
 ASSIGNEE: Chiron S.p.A., Siena, Italy
PATENT INFORMATION: US 6090611 July 18, 2000
SOURCE: Official Gazette of the United States Patent and
 Trademark Office Patents, (July 18, 2000) Vol. 1236,
 No. 3, pp. No Pagination. e-file.
 ISSN: 0098-1133.
DOCUMENT TYPE: Patent
LANGUAGE: English

AB A **cytotoxin associated immunodominant**
 antigen and the nucleic acid encoding the antigen from *Helicobacter*
 pylori are described. This antigen was identified from the
 cytotoxin positive CCUG 17874 *Helicobacter pylori* strain,
 and both the antigen and the DNA encoding it have been sequenced.
 The antigen is a hydrophilic, surface-exposed protein having a
 molecular weight of 120-132 kDa. The nucleic acid encoding the
 antigen may be incorporated into a vector for transformation of host
 cells for expression of the antigen. Both the DNA and the antigen
 can be used in assays for detection of disease or infection by
 Helicobacter pylori, and may find use in treating and
 preventing infection by *Helicobacter pylori* and the
 diseases associated with such infection.

L5 ANSWER 3 OF 7 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:110167 BIOSIS
DOCUMENT NUMBER: PREV200100110167
TITLE: *Helicobacter pylori* proteins useful for
 vaccines and diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford,
 John; Macchia, Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
 ASSIGNEE: Chiron Corporation
PATENT INFORMATION: US 6077706 June 20, 2000
SOURCE: Official Gazette of the United States Patent and
 Trademark Office Patents, (June 20, 2000) Vol. 1235,
 No. 3, pp. No Pagination. e-file.
 ISSN: 0098-1133.

DOCUMENT TYPE: Patent
LANGUAGE: English
AB *Helicobacter pylori* known to cause or be a cofactor in
 type B gastritis, peptic ulcers, and gastric tumors. In both
 developed and developing countries, a high percentage of people are
 infected with this bacterium. The present invention relates
 generally to certain *H. pylori* proteins, to the genes
 which express these proteins, and to the use of these proteins for
 diagnostic and vaccine applications. Specifically, molecular
 cloning, nucleotide, and amino acid sequences for the *H.*
 pylori cytotoxin (CT), the "**Cytotoxin**
 Associated Immunodominant" (CAI)
 antigen, and the heat shock protein (hsp60) are described herein.

L5 ANSWER 4 OF 7 WPIDS COPYRIGHT 2002 DERWENT INFORMATION LTD
ACCESSION NUMBER: 2001-049948 [06] WPIDS
DOC. NO. NON-CPI: N2001-038270
DOC. NO. CPI: C2001-013762

Searcher : Shears 308-4994

THIS PAGE BLANK (USPTO)

09/360685

TITLE: Preparing an agent for diagnosis or control of microbial infection, useful particularly against Helicobacter, based on identification of essential genes in defective mutants.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): APFEL, H; FUCHS, T M; GIBBS, C P; HUECK, C J; MEYER, T F
PATENT ASSIGNEE(S): (CREA-N) CREATOGEN GMBH; (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN
COUNTRY COUNT: 93
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 2000073502	A2	20001207	(200106)*	GE	376
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC					
MW MZ NL OA PT SD SE SL SZ TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK					
DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP					
KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL					
PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU					
ZA ZW					
DE 19924965	A1	20001207	(200106)		
DE 19927740	A1	20001221	(200106)		
DE 19934029	A1	20010125	(200107)		
AU 2000053998	A	20001218	(200118)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE

WO 2000073502	A2	WO 2000-EP5024	20000531
DE 19924965	A1	DE 1999-19924965	19990531
DE 19927740	A1	DE 1999-19927740	19990617
DE 19934029	A1	DE 1999-19934029	19990721
AU 2000053998	A	AU 2000-53998	20000531

FILING DETAILS:

PATENT NO	KIND	PATENT NO

AU 2000053998	A Based on	WO 200073502

PRIORITY APPLN. INFO: DE 1999-19934029 19990721; DE 1999-19924965 19990531; DE 1999-19927740 19990617

AN 2001-049948 [06] WPIDS

AB WO 200073502 A UPAB: 20010126

NOVELTY - Preparation of an agent (A) for detection, prevention and/or treatment of microbial infection by:

(i) identifying essential genes (I) and corresponding polypeptides (II);

(ii) identifying compounds that are directed against (II) and inactivate the microbe;

(iii) testing these for suitability for use; and

(iv) formulating selected (A).

DETAILED DESCRIPTION - Identifying essential genes (I) comprises preparation of gene-deficient microorganisms by conditional antisense inhibition (CAI) and/or subtractive

THIS PAGE BLANK (USPTO)

recombination mutagenesis (SRM), then determining viability and/or survival of the deficient organisms.

INDEPENDENT CLAIMS are also included for the following:

- (a) method for identifying essential microbial genes;
- (b) nucleic acid (Ia) representing essential secretory genes of *Helicobacter*, identified by method (a);
- (c) a gene bank containing at least two (Ia);
- (d) a vector containing at least one (Ia) or its fragment;
- (e) cells containing (Ia) or the vector of (d);
- (f) a mutant bank containing at least two microorganisms transformed with the vectors of (d);
- (g) a polypeptide (IIa) encoded by (Ia) and their immunogenic fragments;
- (h) inhibitors (A') that bind specifically to (IIa), or its fragments, and/or modulates its expression, presentation and/or native function;
- (i) method for preparing (IIa) or its fragments;
- (j) antibodies (Ab), or their fragments, that are specific for (IIa); and
- (k) a pharmaceutical composition containing one of (Ia), the vector of (d), the cells of (e), (IIa), Ab or its fragments, or (A).

ACTIVITY - Antibacterial.

MECHANISM OF ACTION - Passive or active immunization.

USE - (A) (which may be a nucleic acid (Ia), vector or host cell containing (Ia), derived polypeptide (IIa), or fragments, (IIa)-specific antibodies or their fragments or an inhibitor of (IIa)) are particularly used for diagnosis, treatment or prevention of infection by *Helicobacter pylori*. Particularly (Ia) and (IIa) are used in DNA, subunit or live vaccines.

ADVANTAGE - The method identifies essential genes, including those that have homologs in other species, so identified (A) should have a broad spectrum of activity. Many gene-deficient cells can be screened quickly, in an automated process, and the identified genes can be used for screening without purification.

Dwg.0/15

L5 ANSWER 5 OF 7 TOXCENTER COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1996:219941 TOXCENTER

COPYRIGHT: Copyright 2002 ACS

DOCUMENT NUMBER: CA12605055630F

TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors

AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello

CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy.

SOURCE: Proc. Natl. Acad. Sci. U. S. A., (1996) Vol. 93, No. 25, pp. 14648-14653.

CODEN: PNASA6. ISSN: 0027-8424.

COUNTRY: ITALY

DOCUMENT TYPE: Journal

FILE SEGMENT: CAPLUS

OTHER SOURCE: CAPLUS 1996:747006

LANGUAGE: English

ENTRY DATE: Entered STN: 20011116

Last Updated on STN: 20011116

THIS PAGE BLANK (USPTO)

AN 1996:219941 TOXCENTER

CP Copyright 2002 ACS

AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains cagA (cag) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, cag is split into a right segment (cagI) and a left segment (cagII) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, cagI and cagII are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the cagI region and the extreme 3' end of the cagII region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (cagE), which is similar to the toxin-secretion gene of *Bordetella pertussis*, *ptlC*, and the transport systems required for plasmid transfer, including the *virB4* gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the cagI genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the cag region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L5 ANSWER 6 OF 7 WPIDS COPYRIGHT 2002 DERWENT INFORMATION LTD
DUPLICATE 1

ACCESSION NUMBER: 1993-303464 [38] WPIDS

CROSS REFERENCE: 1996-485780 [48]

DOC. NO. CPI: C1993-135210

TITLE: Recombinant *Helicobacter pylori* protein
and corresp. gene - is a cytotoxin, antigen or heat
shock protein used for treating and preventing type
B gastritis, gastric ulcers and gastric tumours.

DERWENT CLASS: B04 D16

INVENTOR(S): BUGNOLI, M; COVACCI, A; MACCHIA, G; RAPPUOLI, R;
TELFORD, J

PATENT ASSIGNEE(S): (ISTS) BIOCINE SPA; (CHIR) CHIRON SPA; (CHIR-N)
CHIRON SPA; (RICE-N) IST RICERCHE IMMUNOBIOLOGICHE
SIENA SRL; (CHIR) CHIRON CORP; (BUGN-I) BUGNOLI M;
(COVA-I) COVACCI A; (MACC-I) MACCHIA G; (RAPP-I)
RAPPUOLI R; (TELF-I) TELFORD J; (ISTS) BIOCINE
SCLAVO SPA

COUNTRY COUNT: 43

PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 9318150	A1	19930916	(199338)*	EN	83
RW: AT BE CH DE DK ES FR GB GR IE IT LU MC NL OA PT SE					
W: AT AU BB BG BR CA CH CZ DE DK ES FI GB HU JP KP KR LK LU MG					
MN MW NL NO NZ PL PT RO RU SD SE SK UA US					
AU 9336300	A	19931005	(199405)		
EP 643770	A1	19950322	(199516)	EN	
R: AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE					
JP 07504565	W	19950525	(199529)		
IT 1262895	B	19960722	(199709)		
SG 48353	A1	19980417	(199828)		
US 5928865	A	19990727	(199936)		

THIS PAGE BLANK (USPTO)

09/360685

EP 967279 A1 19991229 (200005) EN
 R: AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE
 US 6077706 A 20000620 (200035)
 US 6090611 A 20000718 (200037)
 US 6130059 A 20001010 (200052)
 JP 2000333686 A 20001205 (200102) 42
 JP 2000350591 A 20001219 (200104) 51

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9318150	A1	WO 1993-EP472	19930302
AU 9336300	A	AU 1993-36300	19930302
EP 643770	A1	EP 1993-905285	19930302
		WO 1993-EP472	19930302
JP 07504565	W	JP 1993-515309	19930302
		WO 1993-EP472	19930302
IT 1262895	B	IT 1992-FI52	19920302
SG 48353	A1	SG 1996-9100	19930302
US 5928865	A Div ex	WO 1993-EP158	19930125
	Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
	CIP of	US 1995-425194	19950420
	CIP of	US 1995-471491	19950606
		US 1995-477451	19950607
EP 967279	A1 Div ex	EP 1993-905285	19930302
		EP 1999-202698	19930302
US 6077706	A Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-470260	19950606
US 6090611	A Div ex	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-471491	19950606
US 6130059	A Cont of	WO 1993-EP472	19930302
	Div ex	US 1994-256848	19941021
		US 1995-466662	19950606
JP 2000333686	A Div ex	JP 1993-515309	19930302
		JP 2000-126696	19930302
JP 2000350591	A Div ex	JP 1993-515309	19930302
		JP 2000-126695	19930302

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 9336300	A Based on	WO 9318150
EP 643770	A1 Based on	WO 9318150
JP 07504565	W Based on	WO 9318150
EP 967279	A1 Div ex	EP 643770

PRIORITY APPLN. INFO: WO 1993-EP158 19930125; IT 1992-FI52
 19920302

AN 1993-303464 [38] WPIDS

CR 1996-485780 [48]

AB WO 9318150 A UPAB: 20010118

A recombinant *Helicobacter pylori* protein (I), deriv. or fragment is new.

THIS PAGE BLANK (USPTO)

11/11/2011 10:11:11 AM
11/11/2011 10:11:11 AM
11/11/2011 10:11:11 AM

Also new are: (a) a vaccine contg. (I); (b) an immunodiagnostic assay, and kit, having at least 1 step involving at least 1 binding partner, (I), opt. labelled or coupled to a solid support; (c) a recombinant polynucleotide encoding (I), a deriv. or fragment; (d) a nucleic acid assay, and kit, and amplification process using (c) as a probe or primer; (e) a vector contg. (c); and (f) a host cell transformed with (e).

(I) is a cytotoxin or a precursor, pref. exhibiting no or reduced toxicity; a **cytotoxin-associated immunodominant** antigen; or a heat shock protein (sequences in specification).

USE/ADVANTAGE - (I) is used to treat, prevent and diagnose H. **pylori** infection, which causes type B gastritis, peptic ulcers and gastric tumours
Dwg.0/5

L5 ANSWER 7 OF 7 TOXLIT
ACCESSION NUMBER: 1994:22745 TOXLIT
DOCUMENT NUMBER: CA-120-002256H
TITLE: Helicobacter **pylori** proteins useful for vaccines and diagnostics.
AUTHOR: Covacci A; Bugnoli M; Telford J; Macchia G; Rappuoli R
SOURCE: (1993). PCT Int. Appl. PATENT NO. 93 18150 09/16/93 (Biocine Sclavo S.p.A.).
PUB. COUNTRY: Italy
DOCUMENT TYPE: Patent
FILE SEGMENT: CA
LANGUAGE: English
OTHER SOURCE: CA 120:2256
ENTRY MONTH: 199405

AB The H. **pylori** genes for cytotoxin, CAI (**cytotoxin-assocd. immunodominant**) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. **pylori**. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The CAI antigen gene was absent in noncytotoxic H. **pylori** strains. The heterogeneity of the CAI antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. **pylori** strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. **pylori** and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

(FILE 'CAPLUS', MEDLINE, BIOSIS, EMBASE, WPIDS, JICST-EPLUS, JAPIO, TOXLIT, TOXCENTER, PHIC, PHIN' ENTERED AT 10:28:44 ON 14 FEB 2002)

L6 359 SEA ABB=ON PLU=ON COVACCI A?/AU
L7 219 SEA ABB=ON PLU=ON BUGNOLI M?/AU
L8 1040 SEA ABB=ON PLU=ON TELFORD J?/AU
L9 228 SEA ABB=ON PLU=ON MACCHIA G?/AU
L10 2220 SEA ABB=ON PLU=ON RAPPUOLI R?/AU
L11 11 SEA ABB=ON PLU=ON L6 AND L7 AND L8 AND L9 AND L10

- Author (S)

Page 1 of 1

THIS PAGE BLANK (USPTO)

09/360685

L12 206 SEA ABB=ON PLU=ON L6 AND (L7 OR L8 OR L9 OR L10)
L13 125 SEA ABB=ON PLU=ON L7 AND (L8 OR L9 OR L10)
L14 322 SEA ABB=ON PLU=ON L8 AND (L9 OR L10)
L15 17 SEA ABB=ON PLU=ON L9 AND L10
L16 3396 SEA ABB=ON PLU=ON L6 OR L7 OR L8 OR L9 OR L10
L17 9 SEA ABB=ON PLU=ON (L12 OR L13 OR L14 OR L16) AND (CAI
OR (CYTOTOXIN OR CYTO TOXIN) (W) (ASSOC? OR ASS#) (W) (IMMUNO
DOMIN? OR IMMUNO DOMIN?))
X18 19 SEA ABB=ON PLU=ON L11 OR L15 OR L17
L19 11 DUP REM L18 (8 DUPLICATES REMOVED)

L19 ANSWER 1 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:239751 BIOSIS
DOCUMENT NUMBER: PREV200100239751
TITLE: Helicobacter pylori proteins useful for vaccines and
diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli,
Massimo; Telford, John; Macchia,
Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Vc.Provenzano, 8, 53100, Siena Italy
PATENT INFORMATION: US 6130059 October 10, 2000
SOURCE: Official Gazette of the United States Patent and
Trademark Office Patents, (Oct. 10, 2000) Vol. 1239,
No. 2, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent
LANGUAGE: English

AB Helicobacter pylori is known to cause or be a cofactor in type B
gastritis, peptic ulcers, and gastric tumors. In both developed and
developing countries, a high percentage of people are infected with
this bacterium. The present invention relates generally to certain
H. pylori proteins, to the genes which express these proteins, and
to the use of these proteins for diagnostic and vaccine
applications. Specifically, molecular cloning, nucleotide, and amino
acid sequences for the H. pylori cytotoxin (CT), the "
Cytotoxin Associated Immunodominant" (CAI) antigen, and the heat shock protein (hsp60). are
described herein.

L19 ANSWER 2 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
ACCESSION NUMBER: 2001:181462 BIOSIS
DOCUMENT NUMBER: PREV200100181462
TITLE: Helicobacter pylori proteins useful for vaccines and
diagnostics.
AUTHOR(S): Covacci, Antonello (1); Bugnoli,
Massimo; Telford, John; Macchia,
Giovanni; Rappuoli, Rino
CORPORATE SOURCE: (1) Siena Italy
ASSIGNEE: Chiron S.p.A., Siena, Italy
PATENT INFORMATION: US 6090611 July 18, 2000
SOURCE: Official Gazette of the United States Patent and
Trademark Office Patents, (July 18, 2000) Vol. 1236,
No. 3, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent
LANGUAGE: English

AB A cytotoxin associated immunodominant
antigen and the nucleic acid encoding the antigen from Helicobacter

THIS PAGE BLANK (USPTO)

pylori are described. This antigen was identified from the cytotoxin positive CCUG 17874 *Helicobacter pylori* strain, and both the antigen and the DNA encoding it have been sequenced. The antigen is a hydrophilic, surface-exposed protein having a molecular weight of 120-132 kDa. The nucleic acid encoding the antigen may be incorporated into a vector for transformation of host cells for expression of the antigen. Both the DNA and the antigen can be used in assays for detection of disease or infection by *Helicobacter pylori*, and may find use in treating and preventing infection by *Helicobacter pylori* and the diseases associated with such infection.

L19 ANSWER 3 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.

ACCESSION NUMBER: 2001:110167 BIOSIS

DOCUMENT NUMBER: PREV200100110167

TITLE: *Helicobacter pylori* proteins useful for vaccines and diagnostics.

AUTHOR(S): Covacci, Antonello (1); Bugnoli, Massimo; Telford, John; Macchia, Giovanni; Rappuoli, Rino

CORPORATE SOURCE: (1) Siena Italy
ASSIGNEE: Chiron Corporation

PATENT INFORMATION: US 6077706 June 20, 2000

SOURCE: Official Gazette of the United States Patent and Trademark Office Patents, (June 20, 2000) Vol. 1235, No. 3, pp. No Pagination. e-file.
ISSN: 0098-1133.

DOCUMENT TYPE: Patent

LANGUAGE: English

AB *Helicobacter pylori* known to cause or be a cofactor in type B gastritis, peptic ulcers, and gastric tumors. In both developed and developing countries, a high percentage of people are infected with this bacterium. The present invention relates generally to certain *H. pylori* proteins, to the genes which express these proteins, and to the use of these proteins for diagnostic and vaccine applications. Specifically, molecular cloning, nucleotide, and amino acid sequences for the *H. pylori* cytotoxin (CT), the "Cytotoxin Associated Immunodominant" (CAI) antigen, and the heat shock protein (hsp60) are described herein.

L19 ANSWER 4 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 1

ACCESSION NUMBER: 1996:747006 CAPLUS

DOCUMENT NUMBER: 126:55630

TITLE: cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors

AUTHOR(S): Censini, Stefano; Lange, Christina; Xiang, Zhaoying; Crabtree, Jean E.; Ghiara, Paolo; Borodovsky, Mark; Rappuoli, Rino; Covacci, Antonello

CORPORATE SOURCE: Immunobiol. Res. Inst. Siena, Chiron Vaccines, Siena, 53100, Italy

SOURCE: Proc. Natl. Acad. Sci. U. S. A. (1996), 93(25), 14648-14653

CODEN: PNASA6; ISSN: 0027-8424

PUBLISHER: National Academy of Sciences

DOCUMENT TYPE: Journal

LANGUAGE: English

THIS PAGE BLANK (USPTO)

AB CagA, a gene that codes for an immunodominant antigen, is present only in *Helicobacter pylori* strains that are assocd. with severe forms of gastroduodenal disease (type I strains). We found that the genetic locus that contains cagA (cag) is part of a 40-kb DNA insertion that likely was acquired horizontally and integrated into the chromosomal glutamate racemase gene. This pathogenicity island is flanked by direct repeats of 31 bp. In some strains, cag is split into a right segment (cagI) and a left segment (cagII) by a novel insertion sequence (IS605). In a minority of *H. pylori* strains, cagI and cagII are sepd. by an intervening chromosomal sequence. Nucleotide sequencing of the 23,508 base pairs that form the cagI region and the extreme 3' end of the cagII region reveals the presence of 19 ORFs that code for proteins predicted to be mostly membrane assocd. with one gene (cagE), which is similar to the toxin-secretion gene of *Bordetella pertussis*, ptlC, and the transport systems required for plasmid transfer, including the virB4 gene of *Agrobacterium tumefaciens*. Transposon inactivation of several of the cagI genes abolishes induction of IL-8 expression in gastric epithelial cell lines. Thus, we believe the cag region may encode a novel *H. pylori* secretion system for the export of virulence determinants.

L19 ANSWER 5 OF 11 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
DUPLICATE 2

ACCESSION NUMBER: 1994:325647 BIOSIS
DOCUMENT NUMBER: PREV199497338647
TITLE: Virulence determinants in *Helicobacter pylori* infection.
AUTHOR(S): Covacci, A. (1); Censini, S. (1);
Telford, J. L. (1); Bugnoli, M. (1)
; Burroni, D. (1); Dell'orco, M.; Ghiara, P. (1);
Xiang, Z. (1); Macchia, G. (1); Papini, E.;
Montecucco, C.; Rappuoli, R. (1)
CORPORATE SOURCE: (1) Immunobiol. Res. Inst. Siena, Via Fiorentina 1,
53100 Siena Italy
SOURCE: Freer, J. [Editor]; Aitken, R. [Editor]; Alouf, J. E.
[Editor]; Boulnois, G. [Editor]. FEMS Symposium,
(1994) No. 73, pp. 43-50. FEMS Symposium; Bacterial
protein toxins.
Publisher: Gustav Fischer Verlag Wollgrasweg 49,
D-7000 Stuttgart, Germany.
Meeting Info.: Sixth European Workshop Stirling,
Scotland, UK June 27-July 2, 1993
ISSN: 0163-9188. ISBN: 3-437-11535-9, 1-56081-385-7.
DOCUMENT TYPE: Book; Conference
LANGUAGE: English

L19 ANSWER 6 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 3
ACCESSION NUMBER: 1994:2256 CAPLUS
DOCUMENT NUMBER: 120:2256
TITLE: *Helicobacter pylori* proteins useful for vaccines
and diagnostics
INVENTOR(S): Covacci, Antonello; Bugnoli,
Massimo; Telford, John;
Macchia, Giovanni; Rappuoli,
Rino
PATENT ASSIGNEE(S): Biocine Sclavo S.p.A., Italy
SOURCE: PCT Int. Appl., 85 pp.

THIS PAGE BLANK (USPTO)

09/360685

CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9318150	A1	19930916	WO 1993-EP472	19930302
W: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US				
RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG				
AU 9336300	A1	19931005	AU 1993-36300	19930302
EP 643770	A1	19950322	EP 1993-905285	19930302
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE				
JP 07504565	T2	19950525	JP 1993-515309	19930302
EP 967279	A1	19991229	EP 1999-202698	19930302
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE				
JP 2000333686	A2	20001205	JP 2000-126696	19930302
JP 2000350591	A2	20001219	JP 2000-126695	19930302
US 6077706	A	20000620	US 1995-470260	19950606
US 6130059	A	20001010	US 1995-466662	19950606

PRIORITY APPLN. INFO.:

IT 1992-FI52	A	19920302
WO 1993-EP158	A	19930125
EP 1993-905285	A3	19930302
JP 1993-515309	A	19930302
WO 1993-EP472	A	19930302
US 1994-256848	B3	19941021

AB The H. pylori genes for cytotoxin, **CAI** (cytotoxin -assocd. immunodominant) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against H. pylori. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic H. pylori strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all H. pylori strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with H. pylori and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

L19 ANSWER 7 OF 11 TOXLIT
 ACCESSION NUMBER: 1994:22745 TOXLIT
 DOCUMENT NUMBER: CA-120-002256H
 TITLE: Helicobacter pylori proteins useful for vaccines and diagnostics.
 AUTHOR: Covacci A; Bugnoli M;
 Telford J; Macchia G; Rappuoli
 R

THIS PAGE BLANK (USPTO)

09/360685

SOURCE: (1993). PCT Int. Appl. PATENT NO. 93 18150 09/16/93
(Biocine Sclavo S.p.A.).
PUB. COUNTRY: Italy
DOCUMENT TYPE: Patent
FILE SEGMENT: CA
LANGUAGE: English
OTHER SOURCE: CA 120:2256
ENTRY MONTH: 199405

AB The *H. pylori* genes for cytotoxin, **CAI** (cytotoxin -assocd. immunodominant) antigen, and heat-shock protein (of the hsp60 family) are cloned and sequenced. The nucleic acids and proteins may be used for diagnosis and the proteins for vaccination against *H. pylori*. The cytotoxin, or CT antigen, gene encoded a 140 kDa protein precursor of a 100 kDa protein with urease-independent vacuolating activity. The **CAI** antigen gene was absent in noncytotoxic *H. pylori* strains. The heterogeneity of the **CAI** antigen appeared to be due to internal duplications in the gene: strain G39 was found to have two identical repeats of a 102 bp sequence within the gene. The heat-shock protein gene was expressed in all *H. pylori* strains tested, both cytotoxic and noncytotoxic. Most sera from patients infected with *H. pylori* and exhibiting gastritis and ulcers contained antibodies to the hsp, but the degree of recognition varied greatly among the patients and the antibody levels did not show any obvious correlation with the type of disease.

L19 ANSWER 8 OF 11 EMBASE COPYRIGHT 2002 ELSEVIER SCI. B.V.
ACCESSION NUMBER: 93176262 EMBASE
DOCUMENT NUMBER: 1993176262.
TITLE: Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer.
AUTHOR: Covacci A.; Censini S.; Bugnoli M.; Petracca R.; Burrone D.; **Macchia G.**; Massone A.; Papini E.; Xiang Z.; Figura N.; **Rappuoli R.**
CORPORATE SOURCE: Immunobiological Res. Inst. Siena, Via Fiorentina 1, 53100 Siena, Italy
SOURCE: Proceedings of the National Academy of Sciences of the United States of America, (1993) 90/12 (5791-5795).
ISSN: 0027-8424 CODEN: PNASA6
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 004 Microbiology
026 Immunology, Serology and Transplantation
048 Gastroenterology
LANGUAGE: English
SUMMARY LANGUAGE: English

AB *Helicobacter pylori* has been associated with gastritis, peptic ulcer, and gastric adenocarcinoma. We report the nucleotide sequence and expression of an immunodominant antigen of *H. pylori* and the immune response to the antigen during disease. The antigen, named CagA (cytotoxin-associated gene A), is a hydrophilic, surface-exposed protein of 128 kDa produced by most clinical isolates. The size of the cagA gene and its protein varies in different strains by a mechanism that involves duplication of regions within the gene. Clinical isolates that do not produce the antigen do not have the gene and are unable to produce an active

THIS PAGE BLANK (USPTO)

09/360685

vacuolating cytotoxin. An ELISA to detect the immune response against a recombinant fragment of this protein detects 75.3% of patients with gastroduodenal diseases and 100% of patients with duodenal ulcer ($P < 0.0005$), suggesting that only bacteria harboring this protein are associated with disease.

L19 ANSWER 9 OF 11 CAPLUS COPYRIGHT 2002 ACS DUPLICATE 4
ACCESSION NUMBER: 1994:6614 CAPLUS
DOCUMENT NUMBER: 120:6614
TITLE: The Hsp60 protein of Helicobacter pylori:
structure and immune response in patients with
gastroduodenal diseases
AUTHOR(S): **Macchia, Giovanni**; Massone, Annalisa;
Burroni, Daniela; Covacci, Antonello; Censini,
Stefano; **Rappuoli, Rino**
CORPORATE SOURCE: IRIS Immunobiol. Res. Inst. Siena, Siena,
I-53100, Italy
SOURCE: Mol. Microbiol. (1993), 9(3), 645-52
CODEN: MOMIEE; ISSN: 0950-382X
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Helicobacter pylori is a human pathogen that has been assocd. with gastritis, peptic ulcer and gastric carcinoma. The role of the direct action of H. pylori virulence factors and of the induction of autoreactive immunity in the development of chronic gastritis has not been clarified yet. Here the authors report the cloning and mol. characterization of a gene of H. pylori coding for a protein of 58 kDa, recognized by sera of patients affected by H. pylori-induced gastroduodenal diseases. This antigen is present in all the H. pylori strains tested and it belongs to the Hsp60 family of heat-shock proteins, with high homol. with other bacterial and eukaryotic proteins of the same family. This class of homologous proteins has been implicated in the induction of autoimmune disorders in different systems. The presence in infected patients of anti-H. pylori Hsp60 antibodies, potentially cross-reacting with the human homolog, and cross-reactivity between human Hsp60 and a rabbit antiserum against H. pylori Hsp60 suggest that a role of this protein in gastroduodenal diseases is possible.

L19 ANSWER 10 OF 11 JAPIO COPYRIGHT 2002 JPO
ACCESSION NUMBER: 2000-350591 JAPIO
TITLE: VACCINE AND HELICOBACTER PYLORI PROTEIN USEFUL
IN DIAGNOSIS
INVENTOR: **COVACCI ANTONELLO; BUGNOLI
MASSIMO; TELFORD JOHN;
MACCHIA GIOVANNI; RAPPUOLI RINO**
PATENT ASSIGNEE(S): CHIRON SPA)
PATENT INFORMATION:

PATENT NO	KIND	DATE	ERA	MAIN IPC
JP 2000350591A		20001219	Heisei	C12N015-09

JP
APPLICATION INFORMATION
ST19N FORMAT: JP1993-126695 19930302
ORIGINAL: JP2000126695 Heisei
PRIORITY APPLN. INFO.: IT1992FI 52 19920302

Searcher : Shears 308-4994

THIS PAGE BLANK (USPTO)

09/360685

PRIORITY APPLN. INFO.: WO1993EP 9300158 19930125
SOURCE: PATENT ABSTRACTS OF JAPAN (CD-ROM), Unexamined
Applications, Vol. 2000

AN 2000-350591 JAPIO

AB PROBLEM TO BE SOLVED: To obtain a novel Helicobacter pylori protein which possesses a specific sequence containing a specific number of successive amino acids, can be linked with an antibody against Helicobacter pylori and is useful in the diagnosis or prophylaxis of the above bacterial infection and as a vaccine for treatment.

SOLUTION: This Helicobacter pylori protein is a novel recombinant polypeptide which comprises a sequence containing at least 8 successive amino acids and is useful in the diagnosis or prophylaxis of the above bacterial infection and as a vaccine for treatment. The sequence containing at least 8 successive amino acids includes at least a part which can be linked with an antibody against Helicobacter pylori in this sequence consisting of at least 8 successive amino acids or in this polypeptide and is obtained from the amino acid sequence of the formula. This recombinant polypeptide is obtained by preparing the above bacteria-derived DNA library, screening out this library by a partial sequence and expressing the resultant gene.

COPYRIGHT: (C)2000, JPO

L19 ANSWER 11 OF 11 JAPIO COPYRIGHT 2002 JPO

ACCESSION NUMBER: 2000-333686 JAPIO
TITLE: VACCINE AND HELICOBACTER PYLORI PROTEIN USEFUL
FOR DIAGNOSIS

INVENTOR: COVACCI ANTONELLO; BUGNOLI
MASSIMO; TELFORD JOHN;
MACCHIA GIOVANNI; RAPPUOLI RINO

PATENT ASSIGNEE(S): CHIRON SPA)

PATENT INFORMATION:

PATENT NO	KIND	DATE	ERA	MAIN IPC
JP 2000333686A		20001205	Heisei	C12N015-09

JP

APPLICATION INFORMATION

ST19N FORMAT: JP1993-126696 19930302
ORIGINAL: JP2000126696 Heisei
PRIORITY APPLN. INFO.: IT1992FI 52 19920302
PRIORITY APPLN. INFO.: WO1993EP 9300158 19930125
SOURCE: PATENT ABSTRACTS OF JAPAN (CD-ROM), Unexamined
Applications, Vol. 2000

AN 2000-333686 JAPIO

AB PROBLEM TO BE SOLVED: To obtain a new recombinant polypeptide having a specific amino acid sequence, consisting of an antigenic protein of Helicobacter pylori, having immunogenicity and useful for a vaccine for the treatment, the diagnosis and the like of infection of the above bacterium.

SOLUTION: This recombinant polypeptide having immunogenicity, and exhibiting no function or substantially a low function as a toxin has the following characteristics: a new recombinant polypeptide containing a continuous sequence of at least 8 amino acids; the continuous sequence of at least 8 amino acids contains at least one site; the site can be bound to an antibody against Helicobacter pylori both in the sequence consisting of the continuous sequence of

THIS PAGE BLANK (USPTO)

09/360685

at least 8 amino acids and in the polypeptide; the continuous sequence of at least 8 amino acids is obtained from the amino acid sequence of the formula; and this polypeptide is useful for a vaccine, the diagnosis or the like of the infection of the above bacterium.

COPYRIGHT: (C)2000,JPO

=> fil hom

FILE 'HOME' ENTERED AT 10:36:40 ON 14 FEB 2002

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 13:02:31 ; Search time 80.64 Seconds
(without alignments)
1053.597 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 WTNETIDQPTQTEAFNPQQ.....RENAHGKINWTKGFGQKS 1147

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5867	100.0	1147	AA1981	CAI antigen. Heli
2	5867	100.0	1338	AA1980	Helicobacter pylori
3	5281	90.0	1181	AA1983	Tag A antigen of H
4	5281	90.0	1181	AA1983	H. pylori taga ant
5	5281	90.0	1181	AA1983	Helicobacter pylori
6	5086	86.7	1178	AA1984	Helicobacter pylori
7	5086	86.7	1183	AA1984	H. pylori cytoplasm
8	4056.5	69.1	859	AA1984	H. pylori cytoplasm
9	4056.5	69.1	859	AA1984	120-128 kilodalton
10	4031.5	68.7	850	AA1984	H. pylori taga ant
11	1591	27.1	318	AA1989	Helicobacter pylori
					Expressed antigen

12	917	15.6	187	20	AA1989	Antigen 1 from clu
13	326.5	5.6	5024	22	AA1989	S. epidermidis ope
14	307	5.2	1254	11	AA1989	Merozoite apical-en
15	307	5.2	1254	18	AA1989	Merozoite apical-en
16	294	5.0	2688	22	AA1989	Human polypeptide
17	291	5.0	2663	22	AA1989	Human polypeptide
18	290	4.9	1979	21	AA1989	Plasmodium falcipa
19	284	4.8	2285	20	AA1989	Bacillus subtilis
20	281.5	4.8	1279	22	AA1989	S. epidermidis ope
21	275.5	4.7	2485	21	AA1989	Plasmodium falcipa
22	257	4.4	2954	20	AA1989	Amino acid sequenc
23	255	4.3	941	19	AA1989	M. catarrhalis str
24	255	4.3	1392	20	AA1989	Restin protein seq
25	253	4.3	1411	17	AA1989	Nucleolar/endosoma
26	252.5	4.3	980	21	AA1989	Plasmodium falcipa
27	249.5	4.3	2295	21	AA1989	Plasmodium falcipa
28	249	4.2	1639	19	AA1989	P. falciparum synt
29	248.5	4.2	1312	18	AA1989	Human RAD50. Homo
30	248.5	4.2	1312	19	AA1989	Human homologue of
31	248	4.2	1095	22	AA1989	S. epidermidis ope
32	247.5	4.2	1972	17	AA1989	Smooth muscle myos
33	246.5	4.2	1822	13	AA1989	Extracellular fact
34	246	4.2	1119	20	AA1989	B. burgdorferi ant
35	246	4.2	1354	18	AA1989	Physiologically ac
36	246	4.2	1354	19	AA1989	A modified Rho tar
37	246	4.2	1354	20	AA1989	Renal cancer assoc
38	245.5	4.2	975	22	AA1989	Human protein sequ
39	243.5	4.2	1087	20	AA1989	B. burgdorferi ant
40	242	4.1	665	21	AA1989	Plasmodium falcipa
41	242	4.1	1427	12	AA1989	Human 160kD mediat
42	241.5	4.1	1654	6	AA1989	Sequence of the pl
43	241.5	4.1	1788	22	AA1989	Human polypeptide
44	240.5	4.1	1886	19	AA1989	Rattus norvegicus
45	238.5	4.1	1714	21	AA1989	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AA1989
ID AA1989 standard; Protein: 1147 AA.

AC AA1989;

DT 17-MAR-1994 (first entry)

DE CAI antigen.

XX Cytotoxin; CT; H. pylori; precursor; vacuolation; cell death;
KW eukaryote; cytotoxin-associated immunodominant antigen; CAI;
KW heat shock protein; hsp: hsp60; type B; gastritis; peptic ulcer;
KW gastric tumours.

XX Helicobacter pylori.

XX Key Location/Qualifiers
FT Region 600..900
FT Region /note= "Hydrophilic region"
FT Region 703..714
FT Region /note= "Repeat region"
FT Region 748..759
FT Region /note= "Repeat region"
FT Region 880..885
FT Region /note= "Asparagine rich region"
FT Region 890..894
FT Region /note= "Repeat region"
FT Region 909..913
FT Region /note= "Repeat region"
FT Peptide 942..956
FT Peptide /note= "Peptide D3"

PN W09318150-A.

KW ulcer; stomach cancer.
XX Chimeric - Helicobacter pylori.
OS Chimeric - Vibrio cholerae.
XX
FH Key Location/Qualifiers
FT Protein 1..1147
FT Protein /label= CagA
FT Protein 1148..1338
FT Protein /label= A2+B_subunits
XX
XX W09844130-A1.
XX
XX
XX PD 08-OCT-1998.
XX
XX PF 31-MAR-1998; 98WO-KR000073.
XX
XX PR 31-MAR-1997; 97KR-0011951.
XX PR 31-MAR-1997; 97KR-0011950.
XX
XX (DAEW-) DAEWOONG PHARM CO LTD.
XX
XX Choi D, Jung H, Kim B, Park M, Shin S, Yu Y;
XX
XX WPI: 1998-568279/48.
XX DR N-PSDB; AAV62461.
XX
XX New chimeric proteins for use against Helicobacter pylori
PT comprising an antigenic protein of H. pylori and A1 and B subunits
PT of Vibrio cholerae toxin, preferably produced by recombinant
PT techniques
XX
XX Claim 8; Page 83; 102pp: English.
XX
XX This polypeptide comprises a protein fusion between the CagA
CC antigen of Helicobacter pylori and the A2 and B subunits of the
CC Vibrio cholerae toxin. A fusion gene (see AAV62461) encoding the
CC polypeptide is also claimed. The invention relates to: chimeric
CC proteins comprising antigenic proteins of H. pylori and A2 and B
CC subunits of V. cholerae toxin; recombinant DNAs encoding such
CC chimeric proteins; recombinant expression vectors; a process for
CC preparing the chimeric proteins using transformed microbial host
CC cells; and to preventative and therapeutic vaccines comprising the
CC chimeric proteins for H. pylori-associated diseases such as
CC gastritis, gastric ulcer, duodenal ulcer and gastric cancer
CC (claimed). The chimeric proteins are designed to have excellent
CC immunogenicity, to be stable in the stomach, to penetrate the
CC mucous membrane of the intestine, and to stimulate production of
CC sIgA. They can additionally be used as active ingredients in
CC diagnostic kits for H. pylori infection, and for production of
XX anti-H. pylori antibody.
XX
XX Sequence 1338 AA;

Query Match 100.0%; Score 5867; DB 19; Length 1338;
Best Local Similarity 100.0%; Pred. No. 1.3e-294;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTNETIDQQTAEAFNQQFINNLOVAFKVDNAVASYDPDQKPIVDKNDNRDNRQAFEG 60
DB 1 mtnetidqgqtcaafnpqgfinniqvaflkvdnavasydpdqgplvdkdndrnrqafeg 60

QY 61 ISQLREEYSNKAIKPTKKNQYFDFINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
DB 61 isqlreeynskaikptknqyfsdfinksndlinkndlidvesstksfqkfgdqryrif 120

QY 121 TSWVSHQNDPSKINTSRINRMENITQPPILDDKEAEFLKSAKQSFAGIITGNQIRTDQ 180
DB 121 tswvshqndpskintsrinrmenitqppilddkeae flksakqsfagiitgnqirt dq 180

QY 181 KFMGVFDES LKRGAEKNGEPTGGDWLDIFLSFTIDKQSSDVKEAINQEPVPHVQDDI 240
XX

DB 181 kfmgvfdeslkerqaeakngeptggdwldiflsfifdkqssdvkeainqepvphvqddi 240
QY 241 ATTTTDIOGLPEARDLDERGNFSKFTLGDWEMLDVEGVADIDPNYKFNOLLHNNALS 300
DB 241 atttdioglpeardlldergnfskftlgdmemldvegadvadidpnkfnqllhnnals 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNAVGYKDOQGNVATIIIVHMKNGSLV 360
DB 301 svlmgshngiepekvslllyggngpgarhdwnatvgykdqggnvatiiivhmknsgslv 360
QY 361 IAGGEGINNPsfylykEDQLTGSORALSQEIQNKIDFMFLAONNAKLNDLSEKEKEK 420
DB 361 iaggeginnp sfylykedqltgsoral s qe i q n k i d f m f l a o n n a k l n d l s e k e k e k 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAfSVSKDTHKSALITEFGNGDLSYTLKDYGKKADKA 480
DB 421 frteikdfqkds k a y l d a l g n d r i a f s v s k d t h k s a l i t e f g n g d l s y t l k d y g k k a d k a 480
QY 481 LDREKNVTLOGSLKHGDGMFVDYSNFKYTNASKNPKNKGVTNGVSHLEVGFNKVAIFNL 540
DB 481 ldreknvtlogslkhgdgmfvdy snfkytnasknpknkgvtn g v t n g v s h l e v g f n k v a i f n l 540
QY 541 PDLNNLAITTSFVRRLNLDKLTTKGLSPQEAANKLIKDFLSSNKELVGTTLNPNKAVADAKN 600
DB 541 pdlnnlaittsfvr r l n l d k l t t k g l s p q e a n k l i k d f l s s n k e l v g t t l n p n k a v a d a k n 600
QY 601 TGNDEVKKAQKDLKSLRKHLEKEVEKLEKSGNKNKMEAKAQAQNSQKDEIFALIN 660
DB 601 tgndevkkaqkdlk s l r k h l e k e v e k l e s g n k n k m e a k a q a n s q k d e i f a l i n 660
QY 661 KEANDARAIAYAOQLKIGIKRELSDKLENVKNLKDFOKSPDFPKNGKNDFSKAEETLK 720
DB 661 kean d a r a i a y a o q l k i g i k r e l s d k l e n v k n l k d f o k s p d f p k n g k n d f s k a e e t l k 720
QY 721 ALKGSVKDLGNPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVADVIINOKV 780
DB 721 al k g s v k d l g n p e w i s k v e n l n a a l n e f k n g k n d f s k v t o a k s d l e n s v a d v i i n o k v 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYOSV 840
DB 781 t d k v d n l n q a v s v a k a t g d f s r v e q a l a d l k n f s k e q l a q q a k n e s l n a r k k s e i y o s v 840
QY 841 KNGVNTLVGNCLSQAEATTISKNFSDIKKELNAKLGNNNNNNNNNNGLKNEPIYAKVNNKK 900
DB 841 k n g v n t l v g n c l s q a e a t t i s k n f s d i k k e l n a k l g n n n n n n n n n g l k n e p i y a k v n n k k 900
QY 901 AGOAAASLEPIYAQVAKKVNAKIDRLNQTASGLVYVGAAGPPLKRHDKVDLDSKVGLSR 960
DB 901 a g o a a s l e e p i y a q v a k k v n a k i d r l n q t a s g l v y v g a a g p p l k r h d k v d l d s k v g l s r 960
QY 961 NQELAQKIDNLNQAVSEAKAGFFGNLEQITDKLKDSTKHNPMLNWSAKKVPASLSAKL 1020
DB 961 n q e l a q k i d n l n q a v s e a k a g f f g n l e q i t d k l k d s t k h n p m l n w s a k k v p a s l s a k l 1020
QY 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWLKLVNDKIVAHNVGSPVLPSEYDKIG 1080
DB 1021 d n y a t n s h i r i n s n i k n g a i n e k a t g m l t q k n p e w l k l v n d k i v a h n v g s p v l p s e y d k i g 1080
QY 1081 FNQKNMKDYSDFSTKLNNAVKDTNSGFTQFLTNASTASYCYCLARENASHGKKNVNT 1140
DB 1081 f n q k n m k d y s d f s t k l n n a v k d t n s g f t o f l t n a s t a s y c y c l a r e n a s h g k k n v n t 1140
QY 1141 KGGFQKS 1147
DB 1141 k g g f q k s 1147

RESULT 3
AAR53269
ID AAR53269 standard; Protein; 1181 AA.
XX AAR53269;
XX

DT 13-DEC-1994 (first entry)
 XX Tag A antigen of Helicobacter pylori.
 DE Tag A; antigen; peptic ulceration; Helicobacter pylori; antibody;
 XX detection; diagnosis; therapy; treatment.
 KW Helicobacter pylori.
 XX OS
 XX WO9409023-A.
 PN 28-APR-1994.
 XX 13-OCT-1993; 93WO-US09782.
 XX 13-OCT-1992; 92US-0959940.
 PR (UYVA-) UNIV VANDERBILT.
 XX Blaser MJ, Cover TL, Tummuru MKR;
 XX WPI: 1994-151235/18.
 DR N-PSDB; AAQ64581.
 XX
 PT DNA coding tag A gene, from Helicobacter pylori - useful for
 PT detecting predisposition to peptic ulceration
 XX Claim 29; Page 68-72; 87pp; English.
 XX Monoclonal antibodies directed against the antigen may be used to
 CC detect tag A antigen presence which is indicative of a predisposition
 CC to peptic ulceration. A ligand e.g. Antibody, specifically reactive
 CC with the tag A antigen can be used to treat peptic ulcers.
 XX Sequence 1181 AA;
 SQ

Query Match 90.0%; Score 5281; DB 15; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOPEAEAFNPQOFINNLOVAFKVDNAVASVDPDKPIVDKNDPNDNQAFEG 60
 DB 1 mnetidqppceafnpqgfnnlqvafkvdnavasvdpdkpivdkndrdrnqafeg 60
 QY 61 ISQLREEYSNKAIAKPTKKNQYFSDFINKNDLINKMLIDVESSTKSFQFGDQRYRIF 120
 DB 61 lsqllreeysnkaiaikptkknqyfsdfinkndlinkmlidvesstksfkgdgyrif 120
 QY 121 TSWVSHQNDPSKINTRSIRNFWMENIQQPILDDKEAEFLKSQSPAGIIGNOIRTDQ 180
 DB 121 tswvshqndpskintrcirntmehtigppipddkkaeflksaksqspagiignqirtdq 180
 QY 181 KFMGVFDESILKEROBAEKNGEPTGGDWLIDFLSFDFDKKQSDVKAEATNQEPVHPVOPDI 240
 DB 181 kfmgvfdesilkerqaekngptggdwidiflsfddkksdvkeainqepivhpvqdi 240
 QY 241 ATTTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
 DB 241 atstthigglpessrdlldergnfskftlgdmemldvegvdmdpnynfngllhnnlts 300
 QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKDOQGNVATTIINVHMKNGSLV 360
 DB 301 svlmshngiepekvsllayngngfgakhdnatvgykdgqgnnvattinvmknsgslv 360
 QY 361 IAGGEGKINNPFSFYLKEDQLTGSORALSQEEIQNKIDFMEFLAQNNAKLDNLSEKKEK 420
 DB 361 iaggeginnpfsfylekdlqtsqalsqeeiqnkidmfeflqnnakldnlsekeke 420
 QY 421 FRTEIKDFQKSKAYLDALGNDRIDAFVSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
 DB 421 frteikdfqskskpyldalgnndriaafvskdktkhsalitefngdlsytlkvmgkkgika 480

QY 481 LDREKNVTLOGSLKHGCVMFVDYSNFKYTNASKNPNKGVGVTVNGVSHLEVGFNKVAFNLL 540
 DB 481 ldreknvtlqgnlkhdgvmfvnysnfkynaskspnkgvgvtngvshleagfskvafnl 540
 QY 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLRDQLSSNKLKCTTLNFNKAVADAKN 600
 DB 541 pldnnlaitsfvrrrdiedklttkglspqeanlkvdflssnkelvgkalnfinkavaean 600
 QY 601 TGNVDEVKKAQKDLKSLKREHLEKEVEKLESNGKNKMEAKAQAQNSQKDEIFALIN 660
 DB 601 tgnvdevkraqdleklkrehlekvaknlesngknkmeakaqaansqkdeifalln 660
 QY 661 KEANDARAIAYAQNKLKIGIKRELSDKLENVNKLKDFKDFDEFKNGKNKDFSKABETLK 720
 DB 661 keanrdaraiayaqnlkgikrelsdkleninkldkfsksfgfngknkdfksaeetlk 720
 QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVDVVIINQKV 780
 DB 721 alkgsvkdlginpewiskvenlnaalnefkngknkdfskvtqakdsdqlensikdviinqkl 780
 QY 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEOLAQAQAKNESLNARKKSEIYQSV 840
 DB 781 tdkvdnlqavsvakiadfsveqaladlknfskeqlaqqaqknesfnv-gkselygsv 840
 QY 841 KNGVNGTLVGNLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLN--EPIYAKVNMK 898
 DB 840 kngvngtlvnglsgieatalaknfsdikkelneknfnfnnnnnnglknnggepiyaqvmk 898
 QY 899 KKAGQAASLEPIYAQVAKKVNAKIDRLNQIA-SGLGVVVGQA----- 939
 DB 900 kktgvaspeepiyaqvakvktkidqlnqaatsfgvggagfplkrhdkvedlskvr 959
 QY 940 -----AGPPLKRHRDKVDLKVGLSRNQLAOKIDNLQAVSEAKAGEFG 984
 DB 940 svspepiyatiddlggsfplkrhdkvdvlksvlsrndeqltqldnlsqavseakagffg 1019
 QY 985 NLEQITDKLDSTKHNPMLWVESAKVPASLSAKLDNYATNSHTRINSNIKNGAINKA 1044
 DB 1020 nleqtidklkdfknpvnlwaesakvpslsakldnyatnshtrinsnlqngaineka 1079
 QY 1045 TGMLTOKNPEWLKLVNDKIVAHNVGSPVSEYDKTGFNOKNMKYSDSKFSTKUNNAV 1104
 DB 1080 tgmtoqnpewlkivndkivahnvgvspseydnigfsqknmkdysdskfstklinnavk 1139
 QY 1105 DTNSGFTQFLNAPFASVYCLARENAEHLKNNVTGSGFQKS 1147
 DB 1140 diksgftqlanafast-gyysmarenaehglnantkgygfdks 1181

RESULT 4

AAR72593
 ID AAR72593 standard; Protein; 1181 AA.
 XX AC AAR72593;
 XX DT 29-SEP-1995 (first entry)
 XX H. pylori tagA antigen.
 XX TagA; antigen; ulcer; diagnosis; vaccine.
 XX Helicobacter pylori.
 OS US5403924-A.
 PN 04-APR-1995.
 XX 13-OCT-1992; 92US-0959940.
 XX 13-OCT-1992; 92US-0959940.
 XX 26-APR-1993; 93US-0053614.
 PA (UYVA-) UNIV VANDERBILT.

XX Blaser MJ, Cover TL, Tummuru MKR;
 XX WPI: 1995-146855/19.
 DR N-PSDB; AA086728.
 XX
 XX New nucleic acid encoding tag A antigen of *Helicobacter pylori* -
 PT used to detect predisposition to peptic ulceration and to produce
 PT protein for use in vaccines, diagnosis etc.
 XX
 XX Disclosure; Column 37-46; 30pp; English.
 PS
 XX The full-length sequence of the tagA gene of *H. pylori* 84-183 (ATCC
 CC 53726) was obtained from overlapping clones isolated from genomic
 CC libraries. The gene encoded a 1181-amino acid TagA antigen
 CC protein.
 XX
 SQ Sequence 1181 AA;
 Query Match 90.0%; Score 5281; DB 16; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;
 QY 1 MTNETIDQQPQTEAFAFPQFQINNLQVAFILKVDNAVASYDPDQKPIVDKNDNRQAFEG 60
 DB 1 mtnetidqqpqtetfaafpqqfnnlqvafilkvdnavasydpdqkpvdkndnrqafeg 60
 QY 61 ISQLEEYSNKAIKNPKNQYFDFNKSNDLNLKNDLIDVESSTKSFQKFGQDQYRIF 120
 DB 61 isqleeyssnkaiKNPKNQYFDFNKSNDLNLKNDLIDVESSTKSFQKFGQDQYRIF 120
 QY 61 isqleeyssnkaiKNPKNQYFDFNKSNDLNLKNDLIDVESSTKSFQKFGQDQYRIF 120
 QY 121 TSWYSHONDESKINTRSIRNFEMENIIOPLLDDKEAEFLKSAKOSFAGIIIGNOIRTDQ 180
 DB 121 tswyshondepksintrsirnfmenniioPLLDDKEAEFLKSAKOSFAGIIIGNOIRTDQ 180
 QY 181 KFMGVFDESILKEROEAKNGEPTGGDWLIDPLSFIFDKKQSSDVKEAINGPVPVHPQDI 240
 DB 181 kfmgvfdesilkerdeakngptggdwldplsfifdkkqssdvkeaingpvpvhpvqdi 240
 QY 241 ATTTTDTQGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLJHNNALS 300
 DB 241 atstthtqglppesrdldergnfskftlgdmeMLDVEGVADIDPNYKFNOLLJHNNALS 300
 QY 301 SVLMGSHNGIEPERVSLLYGGNGPGARHDMNATVGYKDQGNVATIIINVHMKNGSLV 360
 DB 301 svlmgshngieperksvsllyagngpgarhdmnatvgykdqgnvatiinvhmkngslv 360
 QY 361 IAGGEGINNPSFYLKREDQITGTSORALSQEEIQNKIDFMEFLAONNAKLNLSEKEKEK 420
 DB 361 iaggeginnpsfilykredqitgtsoralSQEEIQNKIDFMEFLAONNAKLNLSEKEKEK 420
 QY 421 FRTEKDFQDKSKAYLDALGNDRTAFVSKDTKHSALITFEGNGDLSYTLKDYCKKADKA 480
 DB 421 frtekdqdkSKAYLDALGNDRTAFVSKDTKHSALITFEGNGDLSYTLKDYCKKADKA 480
 QY 481 LDREKNVTLOGSLKHGDMVFDYSNFKYTNASKNPKNKGVTNGVSHLEVGFNKAIFNL 540
 DB 481 ldreknvtlogslkhgdmvfydysnfkytnasknpknkgvtnvgvshleagfsvavfnl 540
 QY 541 PDLNNAITSVRRNLEDKLTGKLSPOEANKLTKDPLSSNNKELVGKTLNENKAVADAKN 600
 DB 541 pdlnnaitsvrrnledkltgklsPOEANKLTKDPLSSNNKELVGKTLNENKAVADAKN 600
 QY 601 TGNVDEVKKAQDKLEKSLRKEHLEKEVEKKLESKSGNKNKMEAKAANSOKDEIFALIN 660
 DB 601 tgnvdevkkaQDKLEKSLRKEHLEKEVEKKLESKSGNKNKMEAKAANSOKDEIFALIN 660
 QY 661 KEANRDAIRAIAQNLKIGIKRELSDKLENVKNLKDQKSFDEFKNGKNKDFSKAEETLK 720
 DB 661 keanrdairaiaqnlkigikreLSdklenvknLKDQKSFDEFKNGKNKDFSKAEETLK 720
 QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780

DB 721 alkgsvkdilginpewiskvenlnaalinfnknknkdfskvtqksdgnslkdvilnqki 780
 QY 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQALQAOQAQKNSLNARKKSEIYOSV 840
 DB 781 tdkvdelinqavsvakiacdsgveqaladlknfskeqalqaaqaknesfnv-gkseiyqsv 839
 QY 841 KNGVNGTLVNGLSQABATTLTKNFSDIKKELNAKLGNFNNNNNGLKN--EPIYAKVKNK 898
 DB 840 kngvngtlvnglsgleatalknfsdikkelnkfnfnnnnnnglknnggepiyaqvnk 899
 QY 899 KKAGQAASLEPIYAQVAKKVNNAKIDRLNOIA-SGLGVVGOA----- 939
 DB 900 kktggvaspeepiyaqvakvktkiddinqaatsgfgvggagfplkrhdkvedlskvgr 959
 QY 940 -----AGFPLKRHRDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFG 984
 DB 960 svspepiyatidldlgsgfplkrhdkvdlskvlsrqelqtqkldnlsqavseakagffg 1019
 QY 985 NLEQFIDKLKOSTKHNPMLNWLWESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINEKA 1044
 DB 1020 nleqtidklkftknpvnlwaesakkvpaslsakldnyatnshtrinsninqaineke 1079
 QY 1045 TGMLTQKNPEWLKLVNDKIVAHNVGSPVLSYDKIGFNQKNMKDYSDSFKEFTKLNNAVK 1104
 DB 1080 tgrterknpewiklvndkivahnvgsplseydnigfsqkmmkdydsdfkfstklnnavk 1139
 QY 1105 DTNSGTFQFTNFASTASYCLARENAEPHGKIKNTYTKGGFQKS 1147
 DB 1140 diksgftqlanaft-gyyismarenaehglnantkggfgks 1181
 RESULT 5
 AAR91307
 ID AAR91307 standard; Protein; 1181 AA.
 XX AAR91307;
 XX 09-SEP-1996 (first entry)
 XX *Helicobacter pylori* TagA 120-128 kD antigen.
 XX Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;
 KW diagnosis; predisposition; antibody; vaccine; infection.
 XX *Helicobacter pylori*.
 XX WO9610639-A2.
 XX 11-APR-1996.
 XX 29-SEP-1995; 95WO-US12669.
 XX 30-SEP-1994; 94US-0316397.
 XX (ORAV-) ORAVAX INC.
 XX (UYVA-) UNIV VANDERBILT.
 XX Blaser MJ, Cover TL, Kleanthous H, Tummuru MKR;
 XX WPI: 1996-209361/21.
 XX N-PSDB; AAT14051.
 XX *Helicobacter pylori* Tag A gene - used to develop prods. for the
 PT diagnosis, treatment and prevention of peptic ulceration and gastric
 PT carcinoma
 XX Claim 4; Page 103-107; 118pp; English.
 XX AAR91307 is a 120-128 kD antigen of *Helicobacter pylori* (HP). The
 CC antigen is designated TagA and it, or its fragments, can be used to
 CC determine a predisposition to peptic ulceration or gastric carcinoma,
 CC conditions caused by or linked to HP infection. TagA and its

CC fragments may also be used for antibody prodn. for use in detection
 CC of tagA in patients suspected of HP infection. TagA antibodies or
 CC other ligands may also be used to treat peptic ulceration or gastric
 CC carcinoma caused by HP infection. TagA or a non-functional TagA
 CC mutant may be used in vaccines for preventing and treating HP
 CC infection.
 XX
 SQ Sequence 1181 AA;

Query Match 90.0%; Score 5281; DB 17; Length 1181;
 Best Local Similarity 88.4%; Pred. No. 1.8e-264;
 Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOPEAFAFNPOQFINNQLQVAFKVDNAVASYDDQKPIVDKNDNRNQAFEG 60
 DB 1 mtnetidqppqteaaafnpgqfinnqlqafkvdnavasyddqkplvdndrnrdnraqafeg 60

QY 61 ISQLREYSNKAIKNPTKKQNFPSFINKSNBLINKNDLIDVESSTKSQKFGDQRYRF 120
 DB 61 isqlreysnkaiknptkknqnfpsfinksnbldnkndlidvessstksqkfgdgryrif 120

QY 121 TSWSHONDPSPKINTRSIENFENIOPILDDKEAEFLKSAKOSFAGIITGNQIRTDQ 180
 DB 121 tswwshondpskinttrsiernfemtioppilddkkaeflksakosfagiitgnoirtdq 180

QY 181 KFMGVFDESLKQRAEKNGEPTGGDWLIDFLSFIDFKKQSSDVKEAINQEPVPHVQPD 240
 DB 181 kfmgvfdeslkerqeaekngptggdwldiflsfidfkqssdvkeainqepvphvpdi 240

QY 241 ATTTDIOGLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
 DB 241 atsthdiglpesrdlldergrnfsftlgdmemldvegadvadidpnynqllihnnals 300

QY 301 SVLMGSHNGIEPEKYSLYGGNGGPGARHDWATVGYKDOQGNVATINVHMKNGSGLV 360
 DB 301 svlmgshngiepekysllyggnggpgarhdwnatvgykdoqgnvatinvhmkngsglv 360

QY 361 IAGGEGKGINNPSFYLYKEDQLTGSORALSQEEIQKIDFMELFQANNKLDNLSEKEKEK 420
 DB 361 iaggekginnpsfylykedqltgsoralseqeieqkqidmfemeflaqnnakidseskekek 420

QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAIVSKKDKTKHSALTETEGNGDLSTYLDYGRKADKA 480
 DB 421 frneikdfqkdkayldalgnndriaivskkdktkhsaltetegnldstlydygrkaka 480

QY 481 LDREKNVTLOGLSKHDGVNMFVDVSNFKVTNASKNKGVTNGVSHLEVGFKNVAIFNL 540
 DB 481 ldreknvtlqgnlkhdgvnmfvnysnfkvtynaskspnkgvtnvgvshleagfsvavfnl 540

QY 541 PDLNLAITSFVRNLEDKLTTKGLSPQEAANKLIRKDFLSSNKKELVGTINFNKAVADAKN 600
 DB 541 pldnlaitsvvrnledklttkglspqeanlkvdflssnkelvgkalnfnkavaeakn 600

QY 601 TCNYDEVKKAQKDLKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
 DB 601 tcnvdevkkaqkdlkslrkehlekevekklesksgnknkmeakagaansokdeifalin 660

QY 661 KEANDARAIAVAONLKGKIRELSDKLENVKNLKDQKSFDFKNGKNKDFSKAEETLK 720
 DB 661 keandaraiaayonlkgkirelsdklenvknldkdfksfngknknkdfskaeetlk 720

QY 721 ALKGSVKDLGINPEWISKVENLNAALNFKNGKNKDFSKVTQAQSDLENSVKDVIINQKV 780
 DB 721 alkgsvkdlginpewiskvenlnaalnfnknknkdfskvtqaksdngensikdviinqki 780

QY 781 TDKVDNLQAVSVAKATGDFSRVQALADLNFNFKSKEQLAQQAQKNESUNAKKSEIYQSV 840
 DB 781 tdkvdnlqavsvakacdfsrveqaladlnfnfskeqlaqqaqknesfnv-gkseiyqsv 840

QY 841 KNGVNGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKN--EPIYAKVYNK 898
 DB 840 kngvngtlvgnglsqaeatllksnfsdikkelnelkfnfnnnnnnglknnggeiyaqvnk 899

QY 899 KKAGQAASLEPIYAQVAKKVNNAKIDRLNQIA-SGLGVVVGQA----- 939
 DB 900 kktgvaspeepiyaqavakvkkidqlnqaatsfgvggagfplkrhdkvedlsvgr 959

QY 940 -----AGFPLKRHDKVDDLSKVGLSRNOELAQKIDNLNOAVSEAKAGFTG 984
 DB 960 svspepiyatidldlggsfplkrhdkvddlskvglsrnqeltqkidlslsqavseakagf 1019

QY 985 NLEQITDKLKDSTKHNPMNLWVESAKKVPASLSAKLDNATNSHNRINSINIKNGAINKA 1044
 DB 1020 nleqitdklkdftknppnlwvaesakkvpsakldnatnshnrinsnigngaineka 1079

QY 1045 TQMLTQKNPEWMLKLVNDKIVAHNVGSVPLSEYDKTGFNOKNMKDYSDSFKESTKLNNAYK 1104
 DB 1080 tqterqknpewlklvndkivahnvgsvglseydnigfsqkmmkdsdsfkstklinnayk 1139

QY 1105 DPNSTGFTQLTWAFSTASYCLARENAEHGINKVNTKGGFQKS 1147
 DB 1140 diksgftqlanafst-gyysmarenaehgiknantkkgfqs 1181

RESULT 6
 AAW20174
 ID AAW20174 standard; protein; 1178 AA.
 AC AAW20174;
 DT 08-JUL-1997 (first entry)
 XX
 DE H. pylori cytoplasmic protein, 16459375.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 XX WO9640893-A1.
 PD 19-DEC-1996.
 XX 06-JUN-1996; 96WO-US09122.
 PF 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX (ASTR) ASTRA AB.
 PA Berglindh OT, Smith D, Mellgaerd BL;
 PI WPI: 1997-052306/05.
 DR N-PSDB; AAT67410.
 DR
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX Claim 61; Pages 390-392; 1481pp; English.
 PS
 XX The present sequence is a Helicobacter pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 35679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide

[illegible]

Qy	942	-----PPLKRHKVDLDSKVLGRNQELAQKIDNLNQAVSEAKGF	982
Db	955	glsanhepiyatiddigppfplkrhdkvdldskvglsreqkltqkldnlnqavseakash	1014
Qy	983	FGNLEQRTIDKLKDSKTHNPWNLWVEGAKKVPASLSAKLDNYATNSHIRINSIKNGAINE	1042
Db	1015	fdnldqmidkllkdstkknvnllyvesakkvptslakldnyatnshtrinsnvknqtine	1074
Qy	1043	KATGMLTQKNPEWLKLVNDKXIVAHNVGSVPLSYBDYKIGFNQKRMKDYSDSFKPFSTKLNA	1102
Db	1075	katgmleqknewklvndkivahnvgsaplsaydklignqknmkdysdfkfstrlsna	1134
Qy	1103	VKDINSQFTQFLNAESTASYCLARENAEHGIKNNVTKGGFOKS	1147
Db	1135	vkdksgfvgfltnifsmgs-yslmkasvehgvkntntkggfdks	1178
RESULT	7		
AAW20884			
ID	AAW20884	standard; Protein; 1183 AA.	
XX	AAW20884;		
XX	XX		
XX	XX		
DT	18-JUL-1997	(first entry)	
XX	XX		
DE	H. pylori cytoplasmic protein, l3eel2016orf74.		
XX	XX		
KW	Cytoplasmic; vaccine; prevention; infection; envelope;		
KW	identification; binding compound; bacterium; life cycle; activator;		
KW	bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;		
KW	diagnosis.		
XX	XX		
OS	Helicobacter pylori.		
XX	XX		
PN	WO9640893-A1.		
XX	XX		
PD	19-DEC-1996.		
XX	XX		
PF	06-JUN-1996; 96WO-US09122.		
XX	XX		
PR	01-APR-1996; 96US-0630405.		
PR	07-JUN-1995; 95US-0487032.		
XX	XX		
PA	(ASTR) ASTRA AB.		
XX	XX		
PI	Berglindh OF, Smith D, Mellgaard BL;		
XX	XX		
DR	WPI: 1997-052306/05.		
DR	N-PSDB; AAT68137.		
XX	XX		
PT	Helicobacter pylori nucleic acid sequences and related		
PT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori		
PT	infection, and to detect Helicobacter		
XX	XX		
PS	Claim 61; Page 1285-87; 1481pp; English.		
XX	XX		
CC	This sequence represents a H. pylori cytoplasmic protein.		
CC	The protein may be used in a vaccine to prevent or treat H. pylori		
CC	infection or to identify H. pylori polypeptide binding compounds,		
CC	useful as potential H. pylori life cycle activators or inhibitors.		
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from		
CC	overlapping contigs generated by mechanically shearing the bacterial		
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides,		
CC	and the predicted coding regions defined by computer evaluation. To		
CC	identify likely H. pylori antigens for vaccine development, the amino		
CC	acid sequences predicted from various ORF were analysed for significant		
CC	homology to other known or exported membrane proteins. Having identified		
CC	and determined the sequences of interest, particular regions can be		
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide		
CC	production, e.g. in E. coli hosts.		
XX	XX		
SO	Sequence 1183 AA.		

Query Match 86.7%; Score 5086; DB 18; Length 1183;
Best Local Similarity 84.1%; Pred. No. 2.1e-254;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;

Qy 1 MTNETIDQOPOTEAAFPQOFINNQLQVAFKVDNNAVASYDDQKPIVDKNDNRDNRQAFEG 60
Db 17 mnealngqptqteafnpgqfnnlqvafikvdnvvasvdpnqkplvdnkdrndrqafek 76

Qy 61 ISQREEYSNKAINKPTKKNQYFSDFINKSNDL INKNDLIDVESSTKSFQFGDQRYRIF 120
Db 77 isqlreefankaiknptkknqyfsfisksndldkndlidtgssiksfqkfgtqryqif 136

Qy 121 TSWVSHQNDPSKINTSRNFMENIIOPIILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 137 mhwshqndpskintqkirkfmenliqppisddkkaeflrsakqafagilignqirsdq 196

Qy 181 KFMGVFDESLEKEROEAKNGE----PTGGDWLIDFLSFDFDKKQSSDVKEAINQEPVPHV 236
Db 197 kfmgvfdeslkergeaekngengpdtgtdwidflsfvfnkkgssdlketlngepvphv 256

Qy 237 QPDIATTTTIDIOGLPEAPDLDDVEGNFSKFTLGDMEMLDVGVADIDPNYKFNOLLTHN 296
Db 257 qpdvatttdiqslppeardlldergrnfskftlgdmmlldvegvdidpnkfnqlllm 316

Qy 297 NALSSVLGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIINVHMKG 356
Db 317 nalssvlmgshngiepekvslllygngpgpearhwnatvgyknqgdnvatlinvhmkn 376

Qy 357 SGLVTAGSEKGINNPSFYLRKEDQLTGQRALSQSEIQNKIDFMEFLAQNNAKLDNLSEK 416
Db 377 sgvltagsekginnpsfylyrkedqltgqralsqseieiknkvdfmeflaqnnaakldnlskk 436

Qy 417 EKEKTEIKDPQDSKAYLDALGNDRIAFYSKDKTISALITERGNDLSVTLKDYCKK 476
Db 437 ekekfeiedfqskayldalgnndhiafvskkkhlalvaeftngelsytlkdyckk 496

Qy 477 ADKALDRKNTVLOGLSKHGDVYDYSNFYTNASKNPNKGVGTNGVSHLEVGNKVA 536
Db 497 adkaldrakntlqgslkhgdvymfydyynfytnaskspdkvgvatngvshleagfskva 556

Qy 537 IFNLPDLNLAITSFVRNLEDKLTTKGLSPOEANKLTKDFLSSNKELVGTGLNPNKAVA 596
Db 557 vfnlplnlnlatsvvrqleedkllakglspqeanlkvdfllssnkkelvgkalnfnkava 616

Qy 597 DAKNTGNTDEVKKAQDKLESLRKREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIF 656
Db 617 eakntgnydevkkaqdklelskkrrerlekdvaknlesksngknkmeaksgansqkdeif 676

Qy 657 ALINKANRDRATAYANLKGIRKRELSKLEENVNKNLKDQKSPDEFKNGKNKDFSKAE 716
Db 677 alinkeanrdaraiayaqnlkgikrelsdkleninkdikfsksdefkngknkdfskae 736

Qy 717 ETLKALGKSVKDLGINPEWISKVENLNALNEFFKNGKKNKDFSKVTQAKSDLENSKVDVII 776
Db 737 etlkalgksvkdliginpewiskenlnalnefnknknkdfskvtqaksdlsensikdvii 796

Qy 777 NQKVTDKVDNLNOAVSVAKTGDFSRVEQALADLKNFSKEOLAOQAOKNESLNARKKSEI 836
Db 797 nqkicdvdnlnqavsvakatydfsgvgeqaladlnfskeqlaqgaqneofntgknsai 856

Qy 837 YQSVKNGVNGTLVGNGLSQAEATTLSKNFSQTKKELNAKLGNFNNNNNNGLKNPEIYAKV 896
Db 857 yqsvkngvngtlvngnlskaeattlsknfsdtkkelnaaklgnfnnnnnnglens----- 910

Qy 897 NKKAGQAASLEPIYQAVAKVNKAKIDRLNQIASGLGVGQAG----- 941
Db 911 -----tepiytqavkvvakidrlqdasglgdvggaasfllkrhdkvdldskv 959

Qy 942 -----FPLKRHDKVDLDSKVLGRNQLAOKIDNLNQAASVSEAKAGF 982
Db 960 glsanhepiyatidldgpgfplkrhdkvdldskvglsreqkitqkldnlnqavseakash 1019

Qy 983 FGNLEQOTIDKLDSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINE 1042
Db 1020 fgnlqgidkldstckknvnllyvesakkvptslsakldnyatnshtrlnsnvkngtine 1079

Qy 1043 KATGMLTQKNPEWLKLVNDKIVAHNVGVSPVLPSEYDKIGFNQKNMKDYSDSPKSTKLNA 1102
Db 1080 katgmiltgknsowlkvlndkivahnvgvaplaysaydkigfnqknmkdysdfkstrlsna 1139

Qy 1103 VKDTSNGFTQTLTNFASTASYCLARENAEHGKNNVNTKGGPOKS 1147
Db 1140 vkdkisgfvqfntlnfsmgs-yslmkasvehgvkntntkkgfqs 1183

RESULT 8

AAR53268
ID AAR53268 standard; Protein; 859 AA.

AC AAR53268;

DT 13-DEC-1994 (first entry)

XX 120-128 kilodalton antigen of Helicobacter pylori.

Tag A: antigen; peptic ulceration; Helicobacter pylori; antibody;
detection; diagnosis.

OS Helicobacter pylori.

PN WO9409023-A.

XX 28-APR-1994.

PF 13-OCT-1993; 93WO-US09782.

PR 13-OCT-1992; 92US-0959940.

PA (UYVA-) UNIV VANDERBILT.

PI Blaser MJ, Cover TL, Tummuru MKR;

DR WPI; 1994-151235/18.

XX N-PSDB; AAQ64580.

PT DNA coding tag A gene, from Helicobacter pylori - useful for
detecting predisposition to peptic ulceration

XX Claim 3; Page 58-61; 87pp; English.

CC The antigenic fragment is a fragment of the tag A antigen of
Helicobacter pylori. Monoclonal antibodies directed against the
antigen may be used to detect tag A antigen presence which is
indicative of a predisposition to peptic ulceration.

XX Sequence 859 AA;

Query Match 69.1%; Score 4056.5; DB 15; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.7e-201;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOPOTEAAFPQOFINNQLQVAFKVDNNAVASYDDQKPIVDKNDNRDNRQAFEG 60

Db 1 mtnetidqoqptqteafnpgqfnnlqvafikvdnavasyddqkplvdnkdrndrqafeg 60

Qy 61 ISQREEYSNKAINKPTKKNQYFSDFINKSNDL INKNDLIDVESSTKSFQFGDQRYRIF 120

Db 61 isqlreeysnkaiknptkknqyfsdfinksndlnkndnlivesstksfkgdgyrifi 120

Qy 121 TSWVSHQNDPSKINTSRNFMENIIOPIILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180

Db 121 tswvshqndpskintcrlrnfmehtciqpdpdkkaeflksakgsfagilignqirtdq 180

Qy 181 KFMGVFDESLEKEROEAKNGEPTGGDWLIDFLSFDFDKKQSSDVKEAINQEPVPHVQPD 240

```
Db 181 kmgvfdeslkerqaeakngptggdwldflsfldkqssdvkeainqepipvqpd 240
Qy 241 ATTTTDIOGLPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 atstchigglpdesrdlldergnfskftlgdmemldvegvdmdpnkfnqllihntls 300
Qy 301 SVLMGSHNGIPEPEKVSLLYGGNGGPGARHDWNAATVGYKDDQGNVATIIIVHMKNGSGLV 360
Db 301 svlmgshdgiepekvslliyagngfgakhdwatvgvkdqgnnvatliinvhmknsglv 360
Qy 361 TAGGEGKGNPSFYLYKEDLTGTSORALSOEIQNKIDFMEFLAONNAKLNLSEKEK 420
Db 361 taggegknpsfylykedqltgsqralsqeeiqnkidfmeflagnnaksldisekekek 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIFAQVSKDTPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 frneikdfqkdkpyldalgnndriaafvskdkpkhsallitefngdlsytlkvmgkqika 480
Qy 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspnkgvgtngvshleagfskvavfnl 540
Qy 541 PDLNNLAITTSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 pdlnnlaitsvrrrdiedkllakglspeanlkvdfissnkelygkalnfknavaeakn 600
Qy 601 TGNDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNMEAKAQAANSQKDEIFALIN 660
Db 601 tgndevkraqdleksiikkrehlekdvaknlesksngknmeakaqansqkdeifalin 660
Qy 661 KEANDARAIAQAQNLKGIKRELSKLENVNKNLDFDKSDFEPKNGKNKDFSKAEETIK 720
Db 661 keanraraiayaqnlkgikrelsdkleninkldkfskfdgfkngknkdfskaeetik 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVLINQKV 780
Db 721 alkgsvkdlginpewiskenlnaalnefkngknkdfskvtqaksdsgensikdvlingki 780
Qy 781 TDKVDNLNQAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQKNESLNARKKSEIYQSV 840
Db 781 tdkvdnlqavsvakiadcfsgveqaladlnkfskeqlaqaqknesfnv-gkseiyqsv 839
Qy 841 KNGVNGTLVGNGLSQAEAT 859
Db 840 kngvngtlvnglsgleat 858

RESULT 9
ID AAR72594 standard; Protein; 859 AA.
XX
AC AAR72594;
XX
DT 29-SEP-1995 (first entry)
XX
DE H. pylori tagA antigen (truncated).
XX
KW TagA; antigen; ulcer; diagnosis; vaccine.
XX
OS Helicobacter pylori.
XX
PN US5403924-A.
XX
PD 04-APR-1995.
XX
PF 13-OCT-1992; 92US-0959940.
XX
PR 13-OCT-1992; 92US-0959940.
XX
PR 26-APR-1993; 93US-0053614.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
```

```
PI Blaser MJ, Cover TL, Tummuru MKR;
XX
DR WPI; 1995-146855/19.
XX N-PSDB; AAQ86728.
XX
PT New nucleic acid encoding tag A antigen of Helicobacter pylori -
PT used to detect predisposition to peptic ulceration and to produce
PT protein for use in vaccines, diagnosis etc.
XX
PS Disclosure: Column 37-46; 30pp: English.
XX
CC The full-length sequence of the tagA gene of H. pylori 84-183 (ATCC
CC 53726) was obtained from overlapping clones isolated from genomic
CC libraries. The gene encoded a 1181-amino acid TagA antigen
CC protein (AAR72593) and a truncated antigen (AAR72594).
XX
SQ Sequence 859 AA;

Query Match 69.1%; Score 4056.5; DB 16; Length 859;
Best Local Similarity 92.7%; Pred. NO. 1.7e-201;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOPTEAAFPNQOFINNLOVAFKVDNAVASYPDQKPLVDKNDNRDNRQAPEG 60
Db 1 mneicidqgqpteaaafnpqgfinnlqvafkvdnavasydpdqkplvdkndnrdrnqafeg 60
Qy 61 ISQREEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVDESSTKSFQKFGDQYRIF 120
Db 61 isqreeysnkaiknptkknqyfsdfinksndlinknldivvesstksfqkfgdqyrif 120
Qy 121 TSWVSHQNDPSKINTRSRNFMENIIQPILDKKAEPKLSAKOSFAGIIGNIRTDQ 180
Db 121 tswvshqndpskintrcirnmehitqppidpdkkeafelsakgsfagiigndirtq 180
Qy 181 KPMGVFDESLEKROEAKNGEPTGGDWLDIFLSFIFDKKQSSDVKEATNOEPVPHVOPDI 240
Db 181 kpmgvfdeslkerqaeakngptggdwldflsfldkqssdvkeainqepipvqpd 240
Qy 241 ATTTTDIOGLPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 atstchigglpdesrdlldergnfskftlgdmemldvegvdmdpnkfnqllihntls 300
Qy 301 SVLMGSHNGIPEPEKVSLLYGGNGGPGARHDWNAATVGYKDDQGNVATIIIVHMKNGSGLV 360
Db 301 svlmgshdgiepekvslliyagngfgakhdwatvgvkdqgnnvatliinvhmknsglv 360
Qy 361 TAGGEGKGNPSFYLYKEDLTGTSORALSOEIQNKIDFMEFLAONNAKLNLSEKEK 420
Db 361 taggegknpsfylykedqltgsqralsqeeiqnkidfmeflagnnaksldisekekek 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIFAQVSKDTPKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 frneikdfqkdkpyldalgnndriaafvskdkpkhsallitefngdlsytlkvmgkqika 480
Qy 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKNPNKGVGTNGVSHLEVGNKVAIFNL 540
Db 481 ldreknvtlqgnlkhdgvmfvnysnfkytnaskspnkgvgtngvshleagfskvavfnl 540
Qy 541 PDLNNLAITTSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 pdlnnlaitsvrrrdiedkllakglspeanlkvdfissnkelygkalnfknavaeakn 600
Qy 601 TGNDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNMEAKAQAANSQKDEIFALIN 660
Db 601 tgndevkraqdleksiikkrehlekdvaknlesksngknmeakaqansqkdeifalin 660
Qy 661 KEANDARAIAQAQNLKGIKRELSKLENVNKNLDFDKSDFEPKNGKNKDFSKAEETIK 720
Db 661 keanraraiayaqnlkgikrelsdkleninkldkfskfdgfkngknkdfskaeetik 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVLINQKV 780
Db 721 alkgsvkdlginpewiskenlnaalnefkngknkdfskvtqaksdsgensikdvlingki 780
```

Db 721 alkgsvkdglginpewiskvenlnaalnefngknkdskvtqaksdgsensikdvlingki 780

Qy 781 TDKVDNLQAVSAVAKATGDSRVEQALADLNKFSKEQLAQQAQKNSLNARKKSEIYQSV 840

Db 781 tdkvdelngavsvakiacdsgveqaladlnkfskeqlaqaqknesfnv-gkseiyqsv 839

Qy 841 KNGVNGTLVGNLSQAEAT 859

Db 840 kngvngtlvnglsgleat 858

RESULT 10

AAR91308

ID AAR91308 standard; Protein; 850 AA.

XX

AC AAR91308;

XX

DT 09-SEP-1996 (first entry)

DE Helicobacter pylori TagA antigen 96 kD antigenic fragment.

XX

KW Antigen; peptic ulcer; chronic gastritis; gastric adenocarcinoma;

KW diagnosis; predisposition; antibody; vaccine; infection.

XX

OS Helicobacter pylori.

XX

PN WO9610639-A2.

XX

PD 11-APR-1996.

XX

PF 29-SEP-1995; 95WO-US12669.

XX

PR 30-SEP-1994; 94US-0316397.

XX

PA (ORAV-) ORAVAX INC.

PA (UYVA-) UNIV VANDERBILT.

XX

PI Blaser MJ, Cover TL, Kleanthous H, Tummuru MKR;

XX

XX WPI; 1996-209361/21.

DR N-PSDB; AAT14052.

XX

Helicobacter pylori Tag A gene - used to develop prods. for the

PT diagnosis, treatment and prevention of peptic ulceration and gastric

PT carcinoma

XX

Claim 5: Page 93-96; 118pp; English.

XX

AAR91308 is a 96 kD antigenic fragment of the Helicobacter pylori (HP)

CC TagA antigen. TagA, or its fragments, can be used to determine a

CC predisposition to peptic ulceration or gastric carcinoma, both

CC conditions caused by or linked to HP infection. TagA and its

CC fragments may also be used for antibody prodn. for use in detection

CC of TagA in patients suspected of HP infection. TagA antibodies or

CC other ligands may also be used to treat peptic ulceration or gastric

CC carcinoma caused by HP infection. TagA or a non-functional TagA

CC mutant may be used in vaccines for preventing and treating HP

CC infection.

XX

Sequence 850 AA;

Query Match 68.7%; Score 4031.5; DB 17; Length 850;

Best Local Similarity 92.8%; Pred. No. 3.3e-200;

Matches 790; Conservative 24; Mismatches 36; Indels 1; Gaps 1;

Qy 1 MTNETIDQPOTEAAFNPOQFNNLQVAFKVDNAVASYDPDQKPIVDKNDNRNQAPEG 60

Db 1 mtnetldqgqpqteafnppqfnnlqvafkvdnavasydpdqkpivdkndrdnraqefg 60

Qy 61 ISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKDLNLDVESSTKSFQFGDQRYRIF 120

Db 61 isqlreeysnkaiknptkknqyfsdfinksnldlinkdnlivvessstksfkgfgdgyrif 120

Qy 121 TSWSHQNDPSKINTRSIRNFEMENIIQPIILDKEKAEFLKSAQSFAGIIIGNQIRTDQ 180

Db 121 tswshqndpskintrcirnfemehitqppipodkkaeflksakgsfagilignqirtdq 180

Qy 181 KPMGVFDESLSKERQPAEKNGEPTGGDWLDIFLSFFIDFKQSSDYKAEALNQEPVPHVQPD 240

Db 181 kfmvgfdeslkergeaeknggptggdwldiflsfifdkqssdvkealinqepplhvqpdi 240

Qy 241 ATTTTDIOGLPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLINNNALS 300

Db 241 atstthdglpessrdllidernfskftlgdmemldvegvadmdpnymfnglllhnmtls 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGGARHDMNATVGYKDDQGNVATIIINVHMKNGSLV 360

Db 301 svlmgshdgiepekvsllyagnggfgakhdmnatvgykddqgnvatiinvhmkngsglv 360

Qy 361 TAGGKGINNPSFYLYKEDQLTGSQRALISQEBIQNKIDPFMEFLAONNAKLNLSEKEKE 420

Db 361 taggekginnpsfylykedqitgsqralisqebiqnkidpfmeflaonnakldslsekeke 420

Qy 421 PRTEIKDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480

Db 421 frnelkdfqkdskeyldalgnrdria fvskkdpkhsallitefnkgdlsylclvmgkkgika 480

Qy 481 LDREKNVTLOGSLKHGDMFVDSYFNKYTNASKNPKNGVGTNGVSHLEVEGKNVAIFNL 540

Db 481 ldreknvtlqgnlkhgdmfvnysnfkytnaskspkngvgtngvshleagfsvkavfnl 540

Qy 541 PDLNNLAITSFVRRNLEDKLTTKGLSPQEBANKLIKDFLSSNNKELVGKTLNFKNAVADAKN 600

Db 541 pldnnlaitsvrrrdledkltakglspqebanklikdfissnnkelvgkalnfnkavaeakn 600

Qy 601 TGNVDEVKKAOKDLEKSLRKRHLEKEVEKKLESKSGNKNKMEAKAQAQNSOKDEIFALIN 660

Db 601 tgnvdevkradkdekslkrhlekdvaknlesksngknkmeakaqaqnsqdeifalin 660

Qy 661 KEANRDARAIAAQNKGKIKRELSDKLENVNKNLDFOKSFDEPKNGKNKDFSKAEETLK 720

Db 661 keanrdaraiaayqnlkgikrelsdkleninkdlkfsktdgfkngknkdfskaeetlk 720

Qy 721 ALKGSVKDLGINPEWISKVENLNALNFKNGKNKDFSKVTQAKSDLENVSKVDVIINOKV 780

Db 721 alkgsvkdglginpewiskvenlnaalnefngknkdskvtqaksdgsensikdvlingki 780

Qy 781 TDKVDNLQAVSAVAKATGDSRVEQALADLNKFSKEQLAQQAQKNSLNARKKSEIYQSV 840

Db 781 tdkvdelngavsvakiacdsgveqaladlnkfskeqlaqaqknesfnv-gkseiyqsv 839

Qy 841 KNGVNGTLVGN 851

Db 840 kngvngtlvgn 850

RESULT 11

AAR89997

ID AAR89997 standard; Protein; 318 AA.

XX

AC AAR89997;

XX

DT 18-FEB-1999 (first entry)

DE Expressed antigen for cluster 20.

XX

KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;

KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX

OS Helicobacter pylori.

XX

PN WO9849314-A2.

XX

PD 05-NOV-1998.

XX

```

PF 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
PR 25-APR-1997; 97US-0045107.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
PI WPI; 1999-009433/01.
DR
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
XX
XX Claim 16; Page 329; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 318 AA;

Query Match 27.1%; Score 1591; DB 20; Length 318;
Best Local Similarity 95.9%; Pred. No. 7.4e-75;
Matches 305; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 84 SDFINKNDLINKNDLIDVESSTKSFQFGDQRYRIFTSWVSHQNDPSKINTRSIRNFM 143
Db 1 sdfidskndlnkndlidvestsksfqfgdrgyqifstswshqkdpksintrsirnfm 60

QY 144 NIIQPIILDREKAEFLKSAQSAGIIGNOIRTDQFMGVFDESLEKROEAEKNGEPT 203
Db 61 niqipildrekeae flksaqsagiig nqirtdqfmgvfdeslekergeaeknggpt 120

QY 204 GDWLDIFLFTDKQSSDVKEA INQEPVPHVQPIATTTTIDIGLPPEARDLLDERGN 263
Db 121 gdwldiflftdkqssdvkeal nqepvphvqp iattdttdiglppeardllder gn 180

QY 264 FSKFTLGDWEMLDVSGVADIDPNYKFNOLLHNNALSSVLMGSHNGIEPEKVSLLYGGNG 323
Db 181 fskftlgdwemldvsgvadidpnykfnqllhnnalssvlmgshngiepekvsllyagng 240

QY 324 GPGARHDNATVGYKDQGNVATTIIVHMKNGSLVITAGEKGINNPSFYLYKEDQLTGT 383
Db 241 gpgarhdnatvgykdqgnvatt iivhmkngslv itagekginnp sfylykedqltg 300

QY 384 SQRALSQEIEIQNKIDFME 401
Db 301 sqralsqeieirnkvdme 318

RESULT 12
AAW89878
ID AAW89878 standard; Protein: 187 AA.
XX
XX AAW89878;
XX
XX 18-FEB-1999 (first entry)
DT
XX Antigen 1 from cluster 18a.
DE
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

```

```

XX OS Helicobacter pylori.
XX PN WO9849314-A2.
XX PD 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
PR 25-APR-1997; 97US-0045107.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAtee CP;
PI WPI; 1999-009433/01.
DR
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
PT long-lasting immune response
XX
XX Claim 1; Page 225-226; 402pp; English.
XX
XX The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori-positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
CC peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
XX Sequence 187 AA;

Query Match 15.6%; Score 917; DB 20; Length 187;
Best Local Similarity 94.7%; Pred. No. 2.3e-40;
Matches 177; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTNETIDQOPQTEAFAFNQQFINNLOVAFKVDNAVSYDPPQKPIVDKNDNRDNRQAPEG 60
Db 1 mtnetidqopqteafanpqf innlqvafkvdnavasydpdqkplvdkndr dnrqatdg 60

QY 61 ISOLREEYSNKAIKNPTKKNOYFSDFINKSNDLINKNDLIDVESSTKSFQFGDQRYRIF 120
Db 61 sqrlreeysnkaiknptkknqyfsdfinesndlnkndlidvgssiks fqtgtqyrif 120

QY 121 TSWVSHQNDPSKINTRSIRNFMENIQQPILDDKEAEFLKSAQSAGIIGNOIRTDQ 180
Db 121 tswvshqndpskint rsirnfmenliqqpild dkeae flksaqsagi iignoirt dq 180

QY 181 KPMGVFD 187
Db 181 kfmgvfd 187

RESULT 13
AAG82935
ID AAG82935 standard; Protein: 5024 AA.
XX
XX AAG82935;
XX
XX 03-SEP-2001 (first entry)
DT
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2964.
DE
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.

```



```

RESULT 14
AA07503
ID AAR07503 standard; protein; 1254 AA.
XX
AC AAR07503;
XX
DT 06-FEB-1991 (first entry)
XX
DE Merozite apical-end-localised protein (MAEP) insert 5.3.
XX
KW Malaria; vaccine; Duffy blood group.
XX
OS Plasmodium vivax.
XX
PN WO9011772-A.
XX
PD 18-OCT-1990.
XX
PF 03-APR-1990; 90WO-US01849.
XX
PR 06-APR-1989; 89US-0334270.
XX
PR 05-APR-1989; 89US-0334041.
XX
PA (UWNY-) NEW YORK UNIV.
XX
PI Barnwell JW, Galinski MR, Wertheimer SP;
XX
DR WPI; 1990-334616/44.
XX
DR N-PSDB; AAQ06328.
XX
PT Malarial apical end merozoite proteins and peptide(s) - used for
PT developing cpds. for treating, preventing and diagnosing malarial
PT infection
XX
PS Claim 2; Fig 1a; 66pp; English.
XX
CC A MAEP compound having a binding affinity for a Duffy blood group
CC antigen of primate red blood cells, is antigenic for the complete
CC protein, and may be used in diagnosis, treatment and vaccination
CC against invasion by P.vivax and P.knowlesi.
XX
SQ Sequence 1254 AA;

Query Match
Best Local Similarity 5.2%; Score 307; DB 11; Length 1254;
Matches 255; Conservative 221; Mismatches 471; Indels 392; Gaps 65;

QY 7 DOQPOTAAFPNPQFNNLQVAFKVDNAVASYPDPQPIVDKDRDRNQAFEGISQURE 66
DB 2 dkekvkdtstfdekk--ksiekayekmgntlke-----lekmd-deknlekeveeaqi 50
QY 67 EYSNKAIRNPTKKNQYFSDFINKSNDLNLKMDLI-----DVESSTKSQKFGD 114
DB 51 qf-----krifdhvnlmdvekskivmekielykdeideikqkneykqgdt 100
QY 115 QRYRFTSWVSHQNDPSKI-----NTRSRNFMENIIQPP-----ILDOK 154
DB 101 snfytyeqnatsqskakleqfiniattkgtstsdqsdineiesikeevhknqlvkvqes 160
QY 155 EKAEPFLKSAQFAGIIGNQIIRT-----DQRFMGVFDLSLK----- 191
DB 161 nsmeemrkqilsmkdllinnsetiakielnntqalg-frenaktlnktdeqlqrva 219
QY 192 ---ERQAEKNGEPTGGMDLIFLSPFKQSSDYK--EAINQEPVPHVQPIATTDTDI 247
DB 220 mleeakahknn-----ldiale---daqidtevskieginre-imnkkdeiksyylsei 268
QY 248 OGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALSSVLMGSH 307
DB 269 k-----eykdktcteisnkrkgdkkiefie-----kfkpneesnsnkvnineinir 319

```

```

QY 308 NGIEPEKVSLLYGGNGPGARHWNATVGYKDOQGNVATINVHMKNKSGLVIAGGEGK 367
DB 320 qylk-----diedaekgastkvelfhkh-----ett 345
QY 368 INNPSFYLYKEDOLTSQORALSOEIQNKIDFMFLAONNAKLDNLSEKEKEKFEKTEKD 427
DB 346 isn-----ifkeseligve-tksqkinkaedimkieerhnsi-----qtqvkg 389
QY 428 FQKDSKAY-----LDALGNDR-----IAFYSKKDKTHSALITFEGN-----GDLSY 468
DB 390 fgenlnklnephnydaedelndndkstnakvlietnlesvkh--lsetlnkqggekily 447
QY 469 T-LKDYGGKADKALDREKNVTLOGSLKHGDMVFDYSNFKYTN-----ASNPNKGVG-- 520
DB 448 skakdimqkatsentaektie-kvkdqsgnyvnylnqittternlieteknrlngidst 506
QY 521 VTN-----GVSHLEVGFNKVAIFNLPLDNLNLAITSFVRRNLEDKLTGKLSPOEANKL 573
DB 507 itnlegalkesgnyelgf-----lekleeeig-----knrkklvdiikkxi-----nst 550
QY 574 IKDFLSSNKKELVGTNLNFNKAVADAKNTGNYDEVKKAQKDLKSLRK-REHLEKEVEKKL 632
DB 551 vgnfsslfnfdloqydfnknindyen-----kmgeiynefegsnlisenlrnsasents 605
QY 633 ESKSGNKNMKAQAANSQKDEIFALINK--PANRDRARAIAQAQNLKGKRELSKLENV 690
DB 606 dynsaktlrlea-----qkekv-nllnkeeeankylrdvkvkesfrfi-fmkesldki 657
QY 691 NKNLKDFDKSPDEPKNGKNKDFSKAEETLKALKGSKVDLGINPEWISKVENIENALNBEFK 750
DB 658 hemik-----keqltvneghgvkqlvenikelvdenndilsdikq-a 698
QY 751 NGKNKDFSKVTQAQSDLENSVKDVI-----INQKVTD----- 782
DB 699 tgnkeeeiqkith--stlknkaktllghvdtssakygikitpelaltellgdakltaqel 756
QY 793 KVDNLNQAV-----SVAKATGDFSRVEQALADLNKFSKEQOLAQQ-----AQKNESLNARKK 833
DB 757 kfesknvvtletemnsknntneld-vhkniqdaykvvaleilshsdeidtdtkqkssklieng 815
QY 834 SEIYQSV-----KNGVNGTLVGNGLSQAEATTLKSNFSDIKK-----ELN 873
DB 816 nqilykvvlinqyknkissikeeavsvkignvskkhselkskicsksydniialekq 875
QY 874 AKLGNFNN-----NNNN-----GLKN--EPIYAKVNKKAG-----QAASL 907
DB 876 telqnlrnsftqektntnsdsklekiktdfesklnalkktlegevnaalkassdnhvhvqsk 935
QY 908 EEPYIAQVA--KKVNAKIDRLNQIASGLGVVQGAAGFPLKRHRDKVDDLSKVLGSRNQELA 965
DB 936 sepvnpaiseiekeetdidslnatdelikkgrtc--evsryklkldtvtkelsddtel- 992
QY 966 OKIDNLNQAVSAKAGFFGNLEQTIDKL-----KDSKHNPNMLWVESAKKVPASL 1016
DB 993 ---intieknvaylayikknedytvgvdltnehfntkqvsnnheptnf--dksnk-----s 1044
QY 1017 SAKLDNYATNSH-----IRNSNIKNGAINEKATGMLTKQNPWELKLVNOKIVAH 1066
DB 1045 seeltkavtdsktiisklgvlienvntemntiessakeialyn-----elknkktsln 1100
QY 1067 NV-----GSVPLSEYDKIGFNQKNMKDYSDSPKFSFTKLNNNAVKDTNSGFTQFUTNAFSTAS 1122
DB 1101 elyqtsnevklqe-----mksnadkyidvskifntvidtqk--snivtn----- 1142
QY 1123 YYCLARENAEHGKKNVNTK 1141
DB 1143 -----qhsinnvdkk 1152

RESULT 15
AAW24575
ID AAW24575 standard; Protein; 1254 AA.
XX

```


Qy 1017 SAKLDNYATNSH-----IRINSNIKCAINEKATGMLTQKNPEWKLIVNDKIVAH 1066
 Db 1045 seeltkavtdsktiisklgvilevnenntiemntiessakeiealyn-----elknkktsln 1100
 Qy 1067 NV---GSVPLSEYDKIGFNQKNMKDYSDSKFSTKLNNAVKDYNSTGFTQPLTNAFSTAS 1122
 Db 1101 eiqtstnevkige-----mksnadkyldvskifntvldtqk--snivtn----- 1142
 Qy 1123 YYCLARENAEHGKKNVNTK 1141
 Db 1143 -----ghsinnvkdk 1152

Search completed: February 13, 2002, 14:00:58
 Job time: 3507 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 13:05:31 ; Search time 62.29 Seconds
(without alignments)
414.373 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 WTNEIDQPTQTEAAFNPOQ.....RENAEHGKKNVTKGFGQKS 1147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5867	100.0	1147	3	US-08-470-260-5
2	5867	100.0	1147	3	US-08-471-491-5
3	5867	100.0	1147	4	US-08-466-662-5
4	5867	100.0	3289	2	US-08-477-451-2
5	5281	90.0	1181	1	US-08-053-614-4
6	5281	90.0	1181	1	US-08-316-397B-4
7	5281	90.0	1181	2	US-09-034-306-4
8	5281	90.0	1181	4	US-09-259-437-4
9	5281	90.0	1181	5	PCT-US93-09782-4
10	4056.5	69.1	859	1	US-08-033-614-2
11	4056.5	69.1	859	1	US-08-316-397B-2
12	4056.5	69.1	859	2	US-09-034-306-2
13	4056.5	69.1	859	4	US-09-259-437-2
14	4056.5	69.1	859	5	PCT-US93-09782-2
15	284	4.8	2285	4	US-09-308-375-2
16	248.5	4.2	1312	2	US-08-592-126-148
17	248.5	4.2	1312	2	US-08-687-080-51
18	246	4.2	1354	3	US-08-685-871-2
19	238.5	4.1	2710	1	US-08-480-604A-6
20	238.5	4.1	2710	2	US-08-405-496A-6
21	238.5	4.1	2710	4	US-08-915-136-6
22	232	4.0	1085	1	US-08-431-080-28
23	232	4.0	1085	2	US-08-938-534-28
24	231	3.9	1535	3	US-08-755-587-185
25	231	3.9	2391	2	US-08-446-855A-2
26	231	3.9	2391	4	US-09-150-741-2
27	230	3.9	1588	5	PCT-US93-07261-11

28	230	3.9	1663	5	PCT-US93-07261-16
29	228	3.9	1600	2	US-08-617-697-10
30	227.5	3.9	2482	1	US-08-328-254-6
31	226.5	3.9	1786	4	US-08-973-462-8
32	226.5	3.9	3248	1	US-08-353-700-1
33	226.5	3.9	3248	5	PCT-US95-16216-1
34	225.5	3.8	885	2	US-08-533-306A-4
35	225.5	3.8	885	2	US-08-742-923A-4
36	221	3.8	1529	2	US-08-728-470-10
37	221	3.8	1529	4	US-08-719-641-10
38	220	3.7	1664	2	US-08-642-846-2
39	216	3.7	816	2	US-08-533-306A-6
40	216	3.7	816	2	US-08-742-923A-6
41	215	3.7	1494	3	US-08-755-587-186
42	213	3.6	1579	3	US-08-755-587-184
43	212	3.6	1346	1	US-08-471-033-23
44	212	3.6	1346	2	US-08-471-044-23
45	212	3.6	1346	2	US-08-463-483A-23

ALIGNMENTS

RESULT 1
US-08-470-260-5
; Sequence 5, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-260-5

Query Match 100.0% Score 5867; DB 3; Length 1147;

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTNETIDQOQTEAAFNPOQFINNLOVAFKVDNAVASYPDQKPIVDKNDNRDNRQAPEG 60
Db 1 MTNETIDQOQTEAAFNPOQFINNLOVAFKVDNAVASYPDQKPIVDKNDNRDNRQAPEG 60

Qy 61 ISQLEEYSNKAINKPTTKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQRYRIF 120
Db 61 ISQLEEYSNKAINKPTTKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQRYRIF 120

Qy 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180

Qy 181 KPMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVHVQPD 240
Db 181 KPMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVHVQPD 240

Qy 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGLV 360

Qy 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQOEIQNKIDFMEFLAONNAKLNLSEKEK 420
Db 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQOEIQNKIDFMEFLAONNAKLNLSEKEK 420

Qy 421 FRTEIKDFQKSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480

Qy 481 LDREKNVTLQGSKLHGDGVWFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLQGSKLHGDGVWFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFKVAIFNL 540

Qy 541 PDNLNLAITSFVRRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTILNFKAVADAKN 600
Db 541 PDNLNLAITSFVRRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTILNFKAVADAKN 600

Qy 601 TCNYDEVKKAQDLSEKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TCNYDEVKKAQDLSEKSLRKEHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660

Qy 661 KEANRDARAIYAQNLAIGIKRELSDKLENVKNLKDPAKSFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLAIGIKRELSDKLENVKNLKDPAKSFDEPKNGKNKDFSKAEETLK 720

Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVITINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSYKDVITINOKV 780

Qy 781 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQAKNESLNARKKSEIYQSV 840

Qy 841 KNGVNTLVGNGLSOAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNTLVGNGLSOAEATTLSKNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 900

Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNGIASGLGVGQAAGFPPLKRHRDKVDLSKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNGIASGLGVGQAAGFPPLKRHRDKVDLSKVGLSR 960

Qy 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQIDKLDKSTKHNPMLNWLWESAKVPASLSAKL 1020
Db 961 NOELAQKIDNLNQAVSEAKAGFFGNLEQIDKLDKSTKHNPMLNWLWESAKVPASLSAKL 1020

Qy 1021 DNYATNSHIRINSNIKNGAINKATGMLTQKNPEWKLVDKIVAHNVGSPFLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINKATGMLTQKNPEWKLVDKIVAHNVGSPFLSEYDKIG 1080

Db 1021 DNYATNSHIRINSNIKNGAINKATGMLTQKNPEWKLVDKIVAHNVGSPFLSEYDKIG 1080
Qy 1081 FNQKNMKYSDSFKEFTKLNNNAVKDTNSGFTOFLTNAFSTASYICLARENAEHGKKNVAT 1140
Db 1081 FNQKNMKYSDSFKEFTKLNNNAVKDTNSGFTOFLTNAFSTASYICLARENAEHGKKNVAT 1140

Qy 1141 KGGFOKS 1147
Db 1141 KGGFOKS 1147

RESULT 2
US-08-471-491-5
; Sequence 5, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471.491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-471-491-5

Query Match 100.0%; Score 5867; DB 3; Length 1147;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTNETIDQOQTEAAFNPOQFINNLOVAFKVDNAVASYPDQKPIVDKNDNRDNRQAPEG 60
Db 1 MTNETIDQOQTEAAFNPOQFINNLOVAFKVDNAVASYPDQKPIVDKNDNRDNRQAPEG 60

Qy 61 ISQLEEYSNKAINKPTTKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQRYRIF 120
Db 61 ISQLEEYSNKAINKPTTKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQRYRIF 120

Qy 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180

Qy 181 KPMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVHVQPD 240
Db 181 KPMGVFDESLEKROEAEKNGEPTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVHVQPD 240

Qy 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTIDQGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKDOQGNVATTINVHMKNKSGLV 360

Qy 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQOEIQNKIDFMEFLAONNAKLNLSEKEK 420
Db 361 TAGGEGKGINNPSFYLYKEDQLTGSORALSQOEIQNKIDFMEFLAONNAKLNLSEKEK 420

Qy 421 FRTEIKDFQKSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKSKAYLDALGNDRIFAFAVSKDTKHSALITEFGNGDLSYTLKDYGKKADKA 480

Qy 481 LDREKNVTLQGSKLHGDGVWFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLQGSKLHGDGVWFVDYSNFKYTNASKPNKGVGTNGVSHLEVGFKVAIFNL 540
```

```
Db 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Qy 541 PDLNLAITSFVRRLNLEDKLTTKGLSQEANKLTKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDLNLAITSFVRRLNLEDKLTTKGLSQEANKLTKDFLSSNKELVGKTLNFKNAVADAKN 600
Qy 601 TGNVDEVKKAQDKLESLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Qy 661 KEANRDARAIYAQNKGKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNKGKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAOKNESLNARKKSEIYOSV 840
Db 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAOKNESLNARKKSEIYOSV 840
Qy 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGQAAGFPLKRHDKVDLDSKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGQAAGFPLKRHDKVDLDSKVGLSR 960
Qy 961 NOELAOKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHPNMLWVESAKKVPASLSAKL 1020
Db 961 NOELAOKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHPNMLWVESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
```

RESULT 3

```
; Sequence 5, Application US/0846662B
; Patent No. 6130059
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0057
; CURRENT APPLICATION NUMBER: US/08/466.662B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-08-466-662-5
```

Query Match 100.0%; Score 5867; DB 4; Length 1147;
Best Local Similarity 100.0%; Pred. No. 0;

```
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTNETIDQOPQTEAAFNPOQFINNLQVAFKVDNAVAYSDPDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOPQTEAAFNPOQFINNLQVAFKVDNAVAYSDPDQKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQREESYNSKAINKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQREESYNSKAINKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Qy 121 TSWYSHQNDPSKINTSRIRNFEMENIQQPILDDKEKAEFLKSAKOSFAGIIGNRIQTDQ 180
Db 121 TSWYSHQNDPSKINTSRIRNFEMENIQQPILDDKEKAEFLKSAKOSFAGIIGNRIQTDQ 180
Qy 181 KFMGVDESLSKERQEAENKGEPTGGDWLIDIFLSFIDFKKQSSDVKEAINEQEPVPHVQPD 240
Db 181 KFMGVDESLSKERQEAENKGEPTGGDWLIDIFLSFIDFKKQSSDVKEAINEQEPVPHVQPD 240
Qy 241 ATTTTDTIOQLPPEARDLLDERGNFSKFTLGDMEMLDVGCVADIDIDPNYKFNQLLIHNALS 300
Db 241 ATTTTDTIOQLPPEARDLLDERGNFSKFTLGDMEMLDVGCVADIDIDPNYKFNQLLIHNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKQQGNNVATTINVMHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKQQGNNVATTINVMHKNKSGLV 360
Qy 361 TAGKEGINNPFYLYKEDQLTGSQALSQEIEIQNKIDFMEFLAQNNAKLNDLSEKEK 420
Db 361 TAGKEGINNPFYLYKEDQLTGSQALSQEIEIQNKIDFMEFLAQNNAKLNDLSEKEK 420
Qy 421 FRTEIKDQKOSKAYLDALGNDRATFVSKKDTKHSALITFEGNGDLSYTLKDYGGKADKA 480
Db 421 FRTEIKDQKOSKAYLDALGNDRATFVSKKDTKHSALITFEGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMFVDFYDNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAIFNL 540
Qy 541 PDLNLAITSFVRRLNLEDKLTTKGLSQEANKLTKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDLNLAITSFVRRLNLEDKLTTKGLSQEANKLTKDFLSSNKELVGKTLNFKNAVADAKN 600
Qy 601 TGNVDEVKKAQDKLESLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Qy 661 KEANRDARAIYAQNKGKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNKGKRELSDKLENVKNLKDQDKSDFDEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAOKNESLNARKKSEIYOSV 840
Db 781 TDKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAOKNESLNARKKSEIYOSV 840
Qy 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGQAAGFPLKRHDKVDLDSKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGQAAGFPLKRHDKVDLDSKVGLSR 960
Qy 961 NOELAOKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHPNMLWVESAKKVPASLSAKL 1020
Db 961 NOELAOKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHPNMLWVESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
Db 1021 DNYATNSHIRINSNIKNGAINEKATGMLTQKNPEWKLVDNKIVAHNVGVSPLSEYDKIG 1080
```

QY 1081 FNQNMKDYSDSPKSTKLNNVAKDPSGFTQFLTNAPSTASYCLARENABHGKKNVT 1140
 Db 1081 FNQNMKDYSDSPKSTKLNNVAKDPSGFTQFLTNAPSTASYCLARENABHGKKNVT 1140
 QY 1141 KGGFQKS 1147
 Db 1141 KGGFQKS 1147
 RESULT 4
 US-08-477-451-2
 ; Sequence 2, Application US/08477451
 ; Patent No. 5928865
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori CagI Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3289 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-451-2

Query Match 100.0%; Score 5867; DB 2; Length 3289;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTNETIDQOQTEAFNPOQFINNLOVAFKVDNAVASYDDQKPIVDKNDNRNRAQFEG 60
 Db 1532 MTNETIDQOQTEAFNPOQFINNLOVAFKVDNAVASYDDQKPIVDKNDNRNRAQFEG 1591
 QY 61 ISQREEYSNKAINKPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQRYRIF 120
 Db 1592 ISQREEYSNKAINKPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQRYRIF 1651
 QY 121 TSWVSHQNDPSKINTSRINRMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180
 Db 1652 TSWVSHQNDPSKINTSRINRMENIQQPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 1711
 QY 181 KFMGVFDESLEKQEAENGPTGGDWLIDFLSFIDFKKQSSDYKAEINQEPVPHVQDPI 240
 Db 1712 KFMGVFDESLEKQEAENGPTGGDWLIDFLSFIDFKKQSSDYKAEINQEPVPHVQDPI 1771
 QY 241 ATTTTIDIGLPPEARDLLDERGNFSKFTLGDMEMLDVGVADIDPNYKFNQLLIHNNALS 300
 Db 241 ATTTTIDIGLPPEARDLLDERGNFSKFTLGDMEMLDVGVADIDPNYKFNQLLIHNNALS 1831

Db 1772 ATTTTIDIGLPPEARDLLDERGNFSKFTLGDMEMLDVGVADIDPNYKFNQLLIHNNALS 1831
 QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKDQGGNNVATIIINVHMKNGSGLV 360
 Db 1832 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKDQGGNNVATIIINVHMKNGSGLV 1891
 QY 361 IAGGEGKINNPSFYLYKEDQLTGSORALSQEIEQNKIDPFMEFLAQNNAKLDNLSEKEKEK 420
 Db 1892 IAGGEGKINNPSFYLYKEDQLTGSORALSQEIEQNKIDPFMEFLAQNNAKLDNLSEKEKEK 1951
 QY 421 FRTEIKDFOKDSKAYLDALGNDRIAFVSKDTHKHSALITEFNGDLSYTLKDYGGKADKA 480
 Db 1952 FRTEIKDFOKDSKAYLDALGNDRIAFVSKDTHKHSALITEFNGDLSYTLKDYGGKADKA 2011
 QY 481 LDREKNVTLQGSGLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGNKVAIFNL 540
 Db 2012 LDREKNVTLQGSGLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGNKVAIFNL 2071
 QY 541 PDNLNLAITSFVRRNLEDKLTGKLSPOBANKLIKDFLSSNKELVGTGLNFNKAADAKN 600
 Db 2072 PDNLNLAITSFVRRNLEDKLTGKLSPOBANKLIKDFLSSNKELVGTGLNFNKAADAKN 2131
 QY 601 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAQNSQDEIFALIN 560
 Db 2132 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAQNSQDEIFALIN 2191
 QY 661 KEANDARAIAYAQNLIKRELSKLENNVKNLKDFOKSPDEFKNGKNKDFSKAEETLK 720
 Db 2192 KEANDARAIAYAQNLIKRELSKLENNVKNLKDFOKSPDEFKNGKNKDFSKAEETLK 2251
 QY 721 ALKGSVKDLGINPEWISKVENI NAALNEFFKNGKNKDFSKVTQAKSDLENSVKDVIINOKV 780
 Db 2252 ALKGSVKDLGINPEWISKVENI NAALNEFFKNGKNKDFSKVTQAKSDLENSVKDVIINOKV 2311
 QY 781 TDKVDNLNQAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQAKNESLNARKKSEIYQSV 840
 Db 2312 TDKVDNLNQAQVSAKATGDFSRVEQALADLNKFSKEQLAQAQAKNESLNARKKSEIYQSV 2371
 QY 841 KNGVNTLVGNLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 900
 Db 2372 KNGVNTLVGNLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKNEPIYAKVNKK 2431
 QY 901 AGAASLEPIYAQVAKVNAKIDRLNQTASGLGVGQAAGPPLKRHKVDLSDSKVGLSR 960
 Db 2432 AGAASLEPIYAQVAKVNAKIDRLNQTASGLGVGQAAGPPLKRHKVDLSDSKVGLSR 2491
 QY 961 NOELAQKIDNLNQAQVSEAKAGFFGNLEOTIDKLDKSTKHNPMLNVAESAKKVPASLSAKL 1020
 Db 2492 NOELAQKIDNLNQAQVSEAKAGFFGNLEOTIDKLDKSTKHNPMLNVAESAKKVPASLSAKL 2551
 QY 1021 DNYATNSHIRINSINIKNGAINEKATGMLTQKNPEWKLVLVNDKIVAHNVGSPVPLSEYDKIG 1080
 Db 2552 DNYATNSHIRINSINIKNGAINEKATGMLTQKNPEWKLVLVNDKIVAHNVGSPVPLSEYDKIG 2611
 QY 1081 FNQNMKDYSDSPKSTKLNNVAKDPSGFTQFLTNAPSTASYCLARENABHGKKNVT 1140
 Db 2612 FNQNMKDYSDSPKSTKLNNVAKDPSGFTQFLTNAPSTASYCLARENABHGKKNVT 2671
 QY 1141 KGGFQKS 1147
 Db 2672 KGGFQKS 2678

RESULT 5
 US-08-053-614-4
 ; Sequence 4, Application US/08053614
 ; Patent No. 5403924
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: TUMMURU, MURALI K. R.
 ; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/053,614
APPLICATION NUMBER: 19930426
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-053-614-4

Query Match 90.0%; Score 5281; DB 1; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

Qy 1 MTNETIDQPPQTEAFAFPQPPQINNQLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQPPQTEAFAFPQPPQINNQLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60

Qy 1 ISQUREEYSKAIKPKTKKNOYFSDFINKSNDLKNLNDLIVESTSKSFQKFGDOORYIF 120
Db 1 ISQUREEYSKAIKPKTKKNOYFSDFINKSNDLKNLNDLIVESTSKSFQKFGDOORYIF 120

Qy 121 TSWVSHQNDPSKINTSRINRMENIOPPIIDDKAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRINRMENIOPPIIDDKAEFLKSAKOSFAGIIGNOIRTDQ 180

Qy 181 KFMGVFDESLEKQEAKEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQEAKEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240

Qy 241 ATTTTIDIGLPEARDLLDERGNSKFTFLGDMELMDVEGVADIDPNKYFNOLLHNNALS 300
Db 241 ATSTTHIQLPESRDLDERGNSKFTFLGDMELMDVEGVADIDPNKYFNOLLHNNALS 300

Qy 301 SVLMSHGNIPEKVSLLYGCNGPGARHWNATVGYKDOGNVATTIYVHMKNGLV 360
Db 301 SVLMSHGNIPEKVSLLYAGNGEGAKHDWNATVGYKDOGNVATTIYVHMKNGLV 360

Qy 361 IAGGEGINNPFFLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLNLSEKEKEK 420
Db 361 IAGGEGINNPFFLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLNLSEKEKEK 420

Qy 421 FRTEIKFQKDSKAYLDALGNDRIFAIVSKDKTQSALITEFGNGDLSYTLKDYKKADKA 480
Db 421 FRNEIKFQKDSKAYLDALGNDRIFAIVSKDKTQSALITEFGNGDLSYTLKDYKKADKA 480

Qy 481 LDREKNTVLOGSLKHGDMFVDYFNFKYTNASKPNKGVGTNGVSHLEVGFNKAIFNL 540
Db 481 LDREKNTVLOGSLKHGDMFVDYFNFKYTNASKPNKGVGTNGVSHLEVGFNKAIFNL 540

Qy 541 PDANNLAITSFVRNLEDKLTTKGLSPOEANKLKDFLSSNKELVGYKTLNFKAVADAKN 600
Db 541 PDANNLAITSFVRNLEDKLTAKGLSPOEANKLKDFLSSNKELVGYKALNFKAIAEAKN 600

Qy 601 TGNDEVKKAQDLEKSLRKRHELEKEVEKLESGKNGKMEAKQAQANSOKDEIFALIN 660
Db 601 TGNDEVKKAQDLEKSLRKRHELEKEVEKLESGKNGKMEAKQAQANSOKDEIFALIN 660

Qy 661 KEANRDARAIYAQNKLKIKRELSDKLENVKNLKDFOKSEDEFKNGKKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNKLKIKRELSDKLENVKNLKDFOKSEDEFKNGKKNKDFSKAEETLK 720

Qy 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKKNKDFSKVTOAKSDLENSKVDLIINOKV 780
Db 721 ALKGSVDLGINPEWISKVENLNAALNEFKNGKKNKDFSKVTOAKSDLENSKVDLIINOKV 780

Qy 781 TDKVDNLNOAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNOAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840

Qy 841 KNGVNGTLVGNLSQAEATTLTKNFSDFIKKELNAKLGNFNNNNNNGLKN--EPIYAKVKN 898
Db 841 KNGVNGTLVGNLSQAEATTLTKNFSDFIKKELNAKLGNFNNNNNNGLKN--EPIYAKVKN 898

Qy 899 KKAGQASLEPIYAQVAKVNAKIDRLNOIA--SGLGVYVGA----- 939
Db 899 KKAGQASLEPIYAQVAKVNAKIDRLNOIA--SGLGVYVGA----- 939

Qy 939 KKTGOVASPEEPIYAQVAKVTKIDQLNQAAATSGFVGQAGFPLKRHKVDLSKVG 959
Db 939 KKTGOVASPEEPIYAQVAKVTKIDQLNQAAATSGFVGQAGFPLKRHKVDLSKVG 959

Qy 940 -----AGFPLKRHKVDLSKVGSRNQLAQKIDNLNOAVSEAKAGFPG 984
Db 940 -----AGFPLKRHKVDLSKVGSRNQLAQKIDNLNOAVSEAKAGFPG 984

Qy 984 SVSPEPIYATIDDLGGSPFLKRHKVDLSKVGSRNQLAQKIDNLNOAVSEAKAGFPG 1019
Db 984 SVSPEPIYATIDDLGGSPFLKRHKVDLSKVGSRNQLAQKIDNLNOAVSEAKAGFPG 1019

Qy 1019 NLEQTIIDKLDSTKHNPMLNWSAKVPASLSAKLDYATNSHTRINSNTKNGAINKA 1044
Db 1019 NLEQTIIDKLDSTKHNPMLNWSAKVPASLSAKLDYATNSHTRINSNTKNGAINKA 1044

Qy 1044 TGMLETKNPEHLKLVNDKIVAHNVGSVPLSEYDKTIGPNOKNMKYSDSEKFKSLNNAVK 1104
Db 1044 TGMLETKNPEHLKLVNDKIVAHNVGSVPLSEYDKTIGPNOKNMKYSDSEKFKSLNNAVK 1104

Qy 1104 DTNSGFTOFLNAPSTASYCLARENAEHGKKNVNTKGGFOKS 1147
Db 1104 DTNSGFTOFLNAPSTASYCLARENAEHGKKNVNTKGGFOKS 1147

Qy 1147 DKSFTOFLNAPST--GYSWARENAEHGKKNVNTKGGFOKS 1181
Db 1147 DKSFTOFLNAPST--GYSWARENAEHGKKNVNTKGGFOKS 1181

RESULT 6
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE Tga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-4

Query Match 90.0%; Score 5281; DB 1; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOQTEAAFNQQFNNLQVAFKVDNVAASYDDPKPIVDKNDNRDNRQAEFG 60
DB 1 MTNETIDQOQTEAAFNQQFNNLQVAFKVDNVAASYDDPKPIVDKNDNRDNRQAEFG 60

QY 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKPGDQRYRIF 120
DB 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKPGDQRYRIF 120

QY 121 TSWVSHONDPSKINTSRNFMENITIOPIIDDKKEAEFLKSASQSFAGIIGNQIRTDQ 180
DB 121 TSWVSHONDPSKINTSRNFMENITIOPIIDDKKEAEFLKSASQSFAGIIGNQIRTDQ 180

QY 181 KPMGVFDESLEKQERAEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240
DB 181 KPMGVFDESLEKQERAEKNGEPTGGDWLIDFLSFIFDKKQSDVKEAINQEPVPHVQPD 240

QY 241 ATTTTIDIGLPEARDLDERGNFSKFTLGDMEMLDVGVADIDPNYFNQLLIHNNALS 300
DB 241 ATTTTIDIGLPEARDLDERGNFSKFTLGDMEMLDVGVADIDPNYFNQLLIHNNALS 300

QY 301 SVLMGSHNGIEPKVSLYLGNGGPGARHWNATVGYKDDQGNVATINVHMKNKSGLV 360
DB 301 SVLMGSHNGIEPKVSLYLGNGGPGARHWNATVGYKDDQGNVATINVHMKNKSGLV 360

QY 361 IAGGEGKINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKEK 420
DB 361 IAGGEGKINNPSFYLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKEK 420

QY 421 FRTEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITERFGNGDLSTYTKDYGKKADKA 480
DB 421 FRNEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITERFGNGDLSTYTKDYGKKADKA 480

QY 481 LDREKNVTLOGSLKHGDMVYDYNFKYTNASKNKGVTNGVSHLEVGFKNKVAIFNL 540
DB 481 LDREKNVTLOGSLKHGDMVYDYNFKYTNASKNKGVTNGVSHLEVGFKNKVAIFNL 540

QY 541 PDLNNLAITSVYRRNLEKLTTKGLSPQEAANKLIKDFLSSNKELVGTGKTNFKNKAVADAKN 600
DB 541 PDLNNLAITSVYRRNLEKLTTKGLSPQEAANKLIKDFLSSNKELVGTGKTNFKNKAVADAKN 600

QY 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
DB 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660

QY 661 KEANDARAIAAQNKGKIKRELSDKLENVANKLKDPKDFDEKNGKNKDFSKAEETLK 720
DB 661 KEANDARAIAAQNKGKIKRELSDKLENVANKLKDPKDFDEKNGKNKDFSKAEETLK 720

QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAQSDLENSVKDVIINQKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAQSDLENSVKDVIINQKV 780

QY 781 TDKVDNLNAQVSVAKATGDFSRVEQALADLNKFSKEQLAQAQAKNESLARKKSEIYOSV 840

DB 781 TDKVDNLNAQVSVAKIACDFSGVEQALADLNKFSKEQLAQAQAKNESFV - GKSEIYQSV 839
QY 841 KNGVNGTLVGNGLSOAEATTLTSKNFSDIKKELNAKLNFNNNNNCLKN - EPIYAKVVK 898
DB 840 KNGVNGTLVGNGLSGTEATALKNFSDIKKELNEFKNFNNNNNGLKNGGEPPIYAQVVK 899
QY 899 KKAGQAASLEEPPIYAQVAKVNAKIDRLNQIA - SGLGVVQQA - ----- 939
DB 900 KKTGVASPEEPPIYAQVAKVTKKIDQLNQAATSGFGVGQAGFPLKRHDKVEDLSKVCR 959
QY 940 -----AGFPLKRHDKVDDLSKVGLSRNQELAQKIDNLAQVSEAKAGFFG 984
DB 960 SVSPEPIYATIDDLGGSFPLKRHDKVDDLSKVGLSRNQELTQKIDNLSQAVSEAKAGFFG 1019
QY 985 NLEQITIDKLDKSTKUNPMLWVESAKKVPASLSAKLDNATNSHIRINSIKNGAINKA 1044
DB 1020 NLEQITIDKLDKSTKUNPMLWVESAKKVPASLSAKLDNATNSHIRINSIKNGAINKA 1079
QY 1045 TGMLTQKPEWLKLVNDKIVAHNVGSPVLPSEYDKIGFNOKNMKDYSDSFKSTKLNNAYK 1104
DB 1080 TGTERQKPEWLKLVNDKIVAHNVGSPVLPSEYDNIIGFSOKNMKDYSDSFKSTKLNNAYK 1139
QY 1105 DTNSGFTQFLTNASTASYCLARENAEHGKIKNVNTKGGFQKS 1147
DB 1140 DIKSGFTQFLANAFST - GYYSMARENAEHGKIKNVNTKGGFQKS 1181

RESULT 7
US-09-034-306-4
Sequence 4, Application US/09034306
Patent No. 5876943
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-4

Query Match 90.0%; Score 5281; DB 2; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

```
Qy 1 MTNETIDQOPQTEAAFPNQQFINNQLVAFKVDNAVSDPQKPIVDKNDNRQAFEG 60
Db 1 MTNETIDQOPQTEAAFPNQQFINNQLVAFKVDNAVSDPQKPIVDKNDNRQAFEG 60
Qy 61 ISQUREYSNKAIKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQUREYSNKAIKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Qy 121 TSWVSHONDPSKINTRISIRNFMEINIOPIIDDKAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTRICIRNFMEHTIOPIIDDKAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSPIFDKKQSSDVKEAINEQPPVPHVOPDI 240
Db 181 KFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSPIFDKKQSSDVKEAINEQPPVPHVOPDI 240
Qy 241 ATTTTIDQGLPPARDLDRGNFSEKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLDLDRGNFSEKFTLGDMEMLDVEGVADMDPNYKFNOLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDMNATVGYKDOGGNNVATLINVHMKNKSGLV 360
Db 301 SVLMGSHDGEIPEKVSULLYAGNGGFGAKHDMNATVGYKDOGGNNVATLINVHMKNKSGLV 360
Qy 361 IAGEKGINNPSFVLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGEKGINNPSFVLYKEDQLTGSORALSQEEIQNKIDFMEFLAONNAKLDNLSEKKEK 420
Qy 421 FRTEIKDFQKSDKAYLDALGNDRIFAQVSKKDTKHSALITBFGNGDLSYTLKDYKQKADKA 480
Db 421 FRNEIKDFQKSDKPYLDALGNDRIFAQVSKKDPKHSALITBFGNGDLSYTLKVMGKKQIKA 480
Qy 481 LDREKNTVLOGSLKHDGMFVDSNFYKTHASKNPNGVGVTCVSHLEGVFNKVAIFNL 540
Db 481 LDREKNTVLOGSLKHDGMFVDSNFYKTHASKNPNGVGVTCVSHLEGVFNKVAIFNL 540
Qy 541 PDNLNLAITSFVRNRLEDKLTGSLPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Db 541 PDNLNLAITSVVRNLEDKLTGSLPOEANKLIKDFLSSNKELVGKTLNFKAVADAKN 600
Qy 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKLESKGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAQKDLKSLRKEHLEKEVEKLESKGNKNKMEAKAQAQNSQKDEIFALIN 660
Qy 661 KEANRDARATAYAQNLKIGIKRELSDKLENVKNLKDQKDFDKNGKNKDFSKAEETLK 720
Db 661 KEANRDARATAYAQNLKIGIKRELSDKLENVKNLKDQKDFDKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDVINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDVINOKV 780
Qy 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNSLNAFKKSEIYQSV 840
Db 781 TDKVDNLNQAVSVAKIACDFSEVQALADLKNFSKEQLAQQAQKNSFNV-GKSEIYQSV 839
Qy 841 KNGVNGTLVNGLSQAQATTLRSNFSDKKELNAKLNFNNNNNNGLNK--EPIYAKVKNK 898
Db 840 KNGVNGTLVNGLSGIEATLAKNFSIDIKRELNEKFNFNNNNNNGLNKNGEPIYAVQNK 899
Qy 899 KKAQOASLEPIYAVQAKVKNNAKIDRLNOIA-SGLGVWGA----- 939
Db 900 KKTGVASPEPIYAVQAKVKNKIDQLNQAATSGFGVGOAGPPLKRBKVEDLSKVGR 959
Qy 940 -----AGFPLKRHRDKVDLDSKVGLSRNQELAQKIDNLNQAVSEAKAGFFG 984
Db 960 SVSPEPIYATIDDLGGSFPLKRRDKVDLDSKVGLSRNQELTKIDNLNQAVSEAKAGFFG 1019
```

```
Qy 985 NLEOTIDKLDKSTKHNPMNLWVESAKKVPASLSAKLDNATNSHIRINSNIKNGAINEKA 1044
Db 1020 NLEOTIDKLDKSTKHNPMNLWVESAKKVPASLSAKLDNATNSHIRINSNIKNGAINEKA 1079
Qy 1045 TGMUTOKNPEWLKLVNDKIVAHNVGVSPLSEYDKIGFNKNKMDYSDSFSTKLNNAVK 1104
Db 1080 TGTERRQKPEWLKLVNDKIVAHNVGVSPLSEYDNIGFSQKNMDYSDSFSTKLNNAVK 1139
Qy 1105 DTNSGFTQFLTNAESTASYCYCLARENAEHGKKNVNTKGGFOKS 1147
Db 1140 DIKSGFTQFLTNAEST-GYYSMARENAEHGKKNVNTKGGFOKS 1181
```

RESULT 8

US-09-259-437-4

: Sequence 4, Application US/09259437

: Patent No. 6153390

: GENERAL INFORMATION:

: APPLICANT: COVER, TIMOTHY L.

: APPLICANT: BLASER, MARTIN J.

: APPLICANT: HARRY KLEANTHOS

: APPLICANT: TUMURU, MURALI K.R.

: TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

: TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA

: NUMBER OF SEQUENCES: 4

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: NEEDLE & ROSENBERG, P.C.

: STREET: 127 Peachtree Street, Suite 1200

: CITY: Atlanta

: STATE: Georgia

: COUNTRY: USA

: ZIP: 30303

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA: US/09/259,437

: APPLICATION NUMBER: US/09/259,437

: FILING DATE:

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US/08/316,397

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: SPRATT, GWENDOLYN D.

: REGISTRATION NUMBER: 36,016

: REFERENCE/DOCKET NUMBER: 2200.030

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 404/688-0770

: TELEFAX: 404/688-9880

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1181 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: US-09-259-437-4

Query Match 90.0%; Score 5281; DB 4; Length 1181;

Best Local Similarity 88.4%; Pred. No. 0;

Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

Qy 1 MTNETIDQOPQTEAAFPNQQFINNQLVAFKVDNAVSDPQKPIVDKNDNRQAFEG 60

Db 1 MTNETIDQOPQTEAAFPNQQFINNQLVAFKVDNAVSDPQKPIVDKNDNRQAFEG 60

Qy 61 ISQUREYSNKAIKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120

Db 61 ISQUREYSNKAIKNPTKKNOYFSDFTINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120

```
QY 121 TSWVSHQNDPSKINTSRIRNFEMNIIOPIIDDKAEFLKSAKQSFAGIIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRIRNFEMNIIOPIIDDKAEFLKSAKQSFAGIIIGNQIRTDQ 180
QY 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKKQSSDVKEAINOEPPVPHVQDPI 240
Db 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKKQSSDVKEAINOEPPVPHVQDPI 240
QY 241 ATTTTIDQGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
QY 361 TAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
Db 361 TAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
QY 421 FRTEIKDFQKSKAYLDALGNDRIFAIVSKDTKHSALITEFGNGDLSYTLKDYKKADKA 480
Db 421 FRNETKDFQKSKAYLDALGNDRIFAIVSKDTKHSALITEFGNGDLSYTLKVMGKKOIK 480
QY 481 LDREKNVTLOGSLKHGDMVFNYSFKYTNASKPNKGVGTNGVSHLEGVGNKVAIENL 540
Db 481 LDREKNVTLOGSLKHGDMVFNYSFKYTNASKPNKGVGTNGVSHLEAGFSKVAVENL 540
QY 541 PDLNNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNNKELVGTNFKNKAADAKN 600
Db 541 PNLNNLAITSVVRDLEDKLTAKGLSPOEANKLYKDFLSSNNKELVGTNFKNKAADAKN 600
QY 601 TGNVDEVKKAOKLEKSLURKREHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKLEKSLURKREHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
QY 661 KEANDARAIYAONLGIKRELSKLELVNKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIYAONLGIKRELSKLELVNKNLKDPEKSFDEPKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINOKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDQNSIKDVIINOKI 780
QY 781 TDKVDNLNQAQVSAKATGDFSRVQALADLKNFSKEQLAQQAQAKNESLARKKSEIYQSV 840
Db 781 TDKVDNLNQAQVSAKATGDFSRVQALADLKNFSKEQLAQQAQAKNESFNV-GKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAQEAATTLKSNFSDTKKELNAKLGNFNNNNNGLKN--EPIYAKVNK 898
Db 840 KNGVNGTLVGNLSQAQEAATTLKSNFSDTKKELNAKLGNFNNNNNGLKNNGGEPYIAQV 899
QY 899 KKAGQAASLEPIYAQVAKKNVAKIDRLNQIA-SGLGVVGOA----- 939
Db 900 KKTQOVASPEPIYAQVAKKVTYKIDQLNQIATSGFGVGQAGPLKRDHDKVEDLSKVGR 959
QY 940 -----AGFPLKRHDVDDLSKVGLSRNQEALQKIDNLNQAQVSEAKAGFFG 984
Db 960 SVSPEPIYATIDDLGGFPLKRDHDKVDDLSKVGLSRNQEALQKIDNLNQAQVSEAKAGFFG 1019
QY 985 NLEQITDKLKDSTKHPNPNLWVESAKVPASLAKLDNATNSHIRNSINIKGAINEKA 1044
Db 1020 NLEQITDKLKDFTKNNPNVNLWESAKVPASLAKLDNATNSHTRNSINIQGAINEKA 1079
QY 1045 TGMLTQKNPEWLKLVNDKIYAHNVGVSPLSEYDKIGFNOKNMKDYSDSEKFSKLNNAV 1104
Db 1080 TGTQKNPEWLKLVNDKIYAHNVGVSPLSEYDNIGFNSQNMKDYSDSEKFSKLNNAV 1139
QY 1105 DTNSGFTQFLTNASTASYICLARENAEHGINKVNTKGGFQKS 1147
Db 1140 DIKSGFTQFLANAFST-GYYSMARENAEHGINKVNTKGGFQKS 1181
```

```
RESULT 9
PCT-US93-09782-4
; Sequence 4, Application PC/TUS9309782
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: TUMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-09782-4
```

```
Query Match 90.0%; Score 5281; DB 5; Length 1181;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1046; Conservative 33; Mismatches 66; Indels 38; Gaps 5;

QY 1 MTNETIDQOQTEAFNPQQFINNLOVAFKVDNAVASYDDQKPIVDKNDNRNQAFEG 60
Db 1 MTNETIDQOQTEAFNPQQFINNLOVAFKVDNAVASYDDQKPIVDKNDNRNQAFEG 60
QY 61 ISQREEYSNAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKGDQYRIF 120
Db 61 ISQREEYSNAIKNPTKKNQYFSDFINKSNLDINKNDLIVVESSTKSFQKGDQYRIF 120
QY 121 TSWVSHQNDPSKINTSRIRNFEMNIIOPIIDDKAEFLKSAKQSFAGIIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRIRNFEMNIIOPIIDDKAEFLKSAKQSFAGIIIGNQIRTDQ 180
QY 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKKQSSDVKEAINOEPPVPHVQDPI 240
Db 181 KFMGVFDESLEKQERAEKNGEPTGGDWLDFLSFTFDKKQSSDVKEAINOEPPVPHVQDPI 240
QY 241 ATTTTIDQGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATSTTHIQGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKDOQGNVATTINVHMKNKSGSLV 360
QY 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
Db 361 IAGGEKGINNPSFYLYKEDQLTGSORALSQBEIQNKIDFMEFLAONNAKLNDLSEKEK 420
```

Qy	421	FRTETIDFQKDSKAYLDALGNDRIFA	VSFKKDTKHSALITFPGNGDLSYTLKDYQKCKADKA	480	
Db	421	FRNEIKDFQKDSKPYLDALGNDRIFA	VSFKDPKHSALITFPGNGDLSYTLKVMGCKQTKA	480	
Qy	481	LDREKXVTTLQSGSLKHGDGVFV	DSNFKYTNASKPNKGVGTNGVSHLEVGFNKVAIFNL	540	
Db	481	LDREKXVTTLQGNLKHGDGVFV	DSNFKYTNASKPNKGVGTNGVSHLEAGFSKVAIFNL	540	
Qy	541	PDNLNLAIATSVRRNLDELT	TKTGLSPOEANKLIKDFLSSNKELVGTLPNFNKAVADAKN	600	
Db	541	PNLNLAIATSVRRDLEDL	LIAKGLSPOEANKLYKDFLSSNKELVGTLPNFNKAVAEAKN	600	
Qy	601	TGNYDEVKKAQKDLEKSLKRHLEKE	VEVKLEKSGGNKMKMEAKAQAANSOKDIFALIN	660	
Db	601	TGNYDEVKKAQKDLEKSLKRHLEK	DAVNAKLEKSGGNKMKMEAKAQAANSOKDIFALIN	660	
Qy	661	KEANRARAIAVAONLUGIKREL	SOKLENVANKIKDFDKSFDEKFNKGNKDKFSKAEETLK	720	
Db	661	KEANRARAIAVAONLUGIKREL	SOKLENIKNLKDLSKDFGKFNKGNKDKFSKAEETLK	720	
Qy	721	ALAGSVKDLGINPEWISKVEN	LNAALENFKNKDKFSKVTQAOKSDLENSKVDVILINOKV	780	
Db	721	ALAGSVKDLGINPEWISKVEN	LNAALENFKNKDKFSKVTQAOKSDQSENSIKOVILINQKI	780	
Qy	781	TDKVDNLNOQAVSVAKATG	DFSRVEQALADLNKFSKEQLAQQAOKNESLNARKKSEIYQSV	840	
Db	781	TDKVDNLNOQAVSVAKATG	DFSGVEQALADLNKFSKEQLAQQAOKNESPNV--GKSEIYQSV	839	
Qy	841	KNGVNGTLVNGLSQAEAT	TLKSNFSDIKKELNAKLGNFNPNNNNNGLKN--EPIYAKVKN	898	
Db	840	KNGVNGTLVNGLSGTEAT	LAKNFSDIKKELNEKFKFNPNNNNNNGLNKGGEPIYAOVKN	899	
Qy	899	KKAGQAASLEEPIYAOVAK	KVNAKIDRLNOIA--SGLGVVVGQA-----	939	
Db	900	KKTGOYASPEEPIYAOVAK	VTYKIDQLNOAATSGFGVGQAGPFLKRHDVADL	959	
Qy	940	-----	AGFPLKRHDKVDL	LSKVGLSRNOQLAOKIDNLNOQAVSEAKAGFFG	984
Db	960	SVSPEPIYATIDDLGG	SFFPLKRHDKVDL	LSKVGLSRNOQLTQKIDNLNOQAVSEAKAGFFG	1019
Qy	985	NLEQOTDKLKDSTKHN	PNMLNVAESAKKYPASLSAKLDNYATNSHTRINSNLKGAINEKA	1044	
Db	1020	NLEQOTDKLKDFTKNN	PNVNLVAESAKKYPASLSAKLDNYATNSHTRINSNTONGAINEKA	1079	
Qy	1045	TGMLTQKNPEWLKLVNDK	IVAHNVGSPVLSBYDKTGFNOKNMKDYSDSFKEFSTKLNNAVK	1104	
Db	1080	TGTERQKNPEWLKLVNDK	IVAHNVGSPVLSBYDNLGFSQKNMKDYSDSFKEFSTKLNNAVK	1139	
Qy	1105	DTWSGGTQFTLNASTASY	CIARENAEHGINKVNTKGGFQKS	1147	
Db	1140	DIKSGGTQFTLANAFT--GY	SWARNAEHGIKNANTKGGFQKS	1181	

Db 661 KEANRDARAIAAQNLIKRELSKLENINKDLKDFSKSFGFKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDQENSIKDVIINQKI 780
QY 781 TKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKQLAQQAOKNESLNARKKSEIYQSV 840
Db 781 TKVDNLNOAVSVAKIACDFSGVEQALADLKNFSKQLAQQAOKNESFNV-CKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858
RESULT 11
US-08-316-397B-2
; Sequence 2, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316.397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-316-397B-2
Query Match 69.1%; Score 4056.5; DB 1; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;
QY 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAFNPQOFTNNLQVAFKVDNAVASYDDQKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEYSNKALKNPTKKNQVFSDFPINKSNLNDLNIDVESSTKSPQKFGDQRYRIF 120
Db 61 ISQLEEYSNKALKNPTKKNQVFSDFPINKSNLNDLNIDVESSTKSPQKFGDQRYRIF 120
QY 121 TSWVSHQNDPSKINTRSIRNFMEHTIQPPIDDKKAEFLKSAQSFAGIIGNQIRTDQ 180

Db 121 TSWVSHQNDPSKINTRSIRNFMEHTIQPPIDDKKAEFLKSAQSFAGIIGNQIRTDQ 180
QY 181 KPMGVFDESILKERQAEKNGEPTGGDWLDFLSPFDKKOSSDVKKAINEQEPVPHVOPDI 240
Db 181 KPMGVFDESILKERQAEKNGEPTGGDWLDFLSPFDKKOSSDVKKAINEQEPVPHVOPDI 240
QY 241 ATTTTIOGLPPARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFQQLLIHNNALS 300
Db 241 ATSTTHIOGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFQQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGARHDWNATVGYKDOQGNNVATIIINHMKNGSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNATVGYKDOQGNNVATIIINHMKNGSGLV 360
QY 361 IAGGEGKINNPSPYLYKEDOLTGSORALSOEETQNKIDMEFLAONNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPSPYLYKEDOLTGSORALSOEETQNKIDMEFLAONNAKLDNLSEKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFNGDLSYTLKDYGKKADKA 480
Db 421 FRNEIKDFQKDSKPYLDALGNDRIAFVSKKDPKHSALITEFNGDLSYTLKVMGKKQIKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNYSNFKYTNASKNPNKGVGTNGVSHLEVGFENKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNYSNFKYTNASKNPNKGVGTNGVSHLEVGFENKVAIFNL 540
QY 541 POLNLAITSFVRRNLEDLTKTGLSPQAEANKLIKDFLSSNKELVGTNLFNKAVADAKN 600
Db 541 PNLNLAITSFVRRNLEDLTKTGLSPQAEANKLIKDFLSSNKELVGTNLFNKAVADAKN 600
QY 601 TGNVDEVKKAQDKLESLKREHLEKVEKKLESKSGNKNMEAKAQAANSOKDEIFALIN 660
Db 601 TGNVDEVKKAQDKLESLKREHLEKVEKKLESKSGNKNMEAKAQAANSOKDEIFALIN 660
QY 661 KEANRDARAIAAQNLIKRELSKLENINKDLKDFSKSFGFKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIAAQNLIKRELSKLENINKDLKDFSKSFGFKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDQENSIKDVIINQKI 780
QY 781 TKVDNLNOAVSVAKATGDFSRVEQALADLKNFSKQLAQQAOKNESLNARKKSEIYQSV 840
Db 781 TKVDNLNOAVSVAKIACDFSGVEQALADLKNFSKQLAQQAOKNESFNV-CKSEIYQSV 839
QY 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858
RESULT 12
US-09-034-306-2
; Sequence 2, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034.306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/316.397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-2

Query Match 69.1%; Score 4056.5; DB 2; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;

Qy 1 MTNETIDQOQTEAAFNPOQFINNQLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAAFNPOQFINNQLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Qy 61 ISQLREEYSKAIKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Db 61 ISQLREEYSKAIKPTKKNQYFSDFINKSNDLINKNDLIDVESSTKSFQKFGDQRYRIF 120
Qy 121 TSWVSHQNDPSKINTSRINFMENIQQPIIDDKKAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRINFMENIQQPIIDDKKAEFLKSAKOSFAGIIGNOIRTDQ 180
Qy 181 KFMGVFDESLEKQEAENGEPTGDMWLDIFLSFIDFKKQSSDVKEAINEQEPVPHVQDPI 240
Db 181 KFMGVFDESLEKQEAENGEPTGDMWLDIFLSFIDFKKQSSDVKEAINEQEPVPHVQDPI 240
Qy 241 ATTTIDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Db 241 ATSTTHIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAATVGYKQDQGNVATTIINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGPGARHDWNAATVGYKQDQGNVATTIINVHMKNGSLV 360
Qy 361 TAGGEGKINNPSFYLYKEDQLTGSORALSQEIQNKIDFMFEFLAONNAKLDNLSKEKEK 420
Db 361 TAGGEGKINNPSFYLYKEDQLTGSORALSQEIQNKIDFMFEFLAONNAKLDNLSKEKEK 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRIAFVSKDQTHSALITFEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRNEIKDFQKDSKAYLDALGNDRIAFVSKDQTHSALITFEFGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMFYVDYSNFKYTNASKNPKNGVGTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMFYVDYSNFKYTNASKNPKNGVGTNGVSHLEVGFNKVAIFNL 540
Qy 541 PDLNNLAITSFVRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PNLNNLAITSFVRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Qy 601 TCNYDEVKAKQDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660
Db 601 TCNYDEVKAKQDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAANSOKDEIFALIN 660

Qy 661 KEANDARAIYAQNKGIGIKRELSDKLENVNKNLKDFOKSDFEFKNGKNKDFSKAETLK 720
Db 661 KEANDARAIYAQNKGIGIKRELSDKLENVNKNLKDFOKSDFEFKNGKNKDFSKAETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVVDVINOQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVVDVINOQKV 780
Qy 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLNKFNSEQLAQQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLNKFNSEQLAQQAQAKNESLNARKKSEIYQSV 840
Qy 841 KNGVNGTLVGNLSQAEAT 859
Db 840 KNGVNGTLVGNLSGIEAT 858
RESULT 13
US-09-259-437-2
; Sequence 2, Application US/09259437
; Patent No. 6153390
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOS
; APPLICANT: TUMMURU, MURALI K. R.
; TITLE OF INVENTION: THE TagA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/259.437
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316.397
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-259-437-2

Query Match 69.1%; Score 4056.5; DB 4; Length 859;
Best Local Similarity 92.7%; Pred. No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;
Qy 1 MTNETIDQOQTEAAFNPOQFINNQLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAAFNPOQFINNQLQVAFKVDNAVASDPPQKPIVDKNDNRDNRQAFEG 60

```
QY 61 ISQREYSNKAINKPTKKNQYFSDFINNKNDLNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 ISQREYSNKAINKPTKKNQYFSDFINNKNDLNDLIDVESSTKSFQKFGDQYRIF 120
QY 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQERBAEKNQEGTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQERBAEKNQEGTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
Db 241 ATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
QY 361 IAGGKGINNPSFYLYKEDQLTGSORALSQEBEQNKIDFMEFLAQNNAKLDLSEKEK 420
Db 361 IAGGKGINNPSFYLYKEDQLTGSORALSQEBEQNKIDFMEFLAQNNAKLDLSEKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
QY 541 PDNNLAITSFVRNLEDKLTQGLSPQEBANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
Db 541 PDNNLAITSFVRNLEDKLTQGLSPQEBANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
QY 601 TGNVDEVKKAOKLEKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKLEKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
QY 661 KEANDARAIAQNLKIGIKRELSDKLENVANKLKDPSKDFPKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIAQNLKIGIKRELSDKLENVANKLKDPSKDFPKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKOLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVADVIINQV 780
Db 721 ALKGSVKOLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVADVIINQV 780
QY 781 TDKVDNLQAVSVAKATGDFSRVEQALADLNFKSEQLAQQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLQAVSVAKATGDFSRVEQALADLNFKSEQLAQQAQAKNESLNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSQAEAT 859
Db 841 KNGVNGTLVGNGLSQAEAT 859
```

RESULT 14

PCT-US93-09782-2

; Sequence 2, Application PC/TUS9309782

; GENERAL INFORMATION:

; APPLICANT: COVER, TIMOTHY L.

; APPLICANT: BLASER, MARTIN J.

; APPLICANT: TUMMURU, MURALI K. R.

; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NEEDLE & ROSENBERG, P.C.

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

```
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09782
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-09782-2
```

```
Query Match 69.1% Score 4056.5 DB 5; Length 859;
Best Local Similarity 92.7% Pred No. 1.1e-249;
Matches 796; Conservative 24; Mismatches 38; Indels 1; Gaps 1;
```

```
QY 1 MTNETIDQPOQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQPOQTEAAFNPOQFINNLOVAFKVDNAVASYDDPKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEESYNSKAINKPTKKNQYFSDFINNKNDLNDLIDVESSTKSFQKFGDQYRIF 120
Db 61 ISQLEESYNSKAINKPTKKNQYFSDFINNKNDLNDLIDVESSTKSFQKFGDQYRIF 120
QY 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
Db 121 TSWVSHONDPSKINTSRINFMENIQQPILDDKEAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQERBAEKNQEGTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVFDESLEKQERBAEKNQEGTGGDWLDIFLSFIFDKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
Db 241 ATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGARHDNATVGYKDOQGNVATINVMHKNKSGLV 360
QY 361 IAGGKGINNPSFYLYKEDQLTGSORALSQEBEQNKIDFMEFLAQNNAKLDLSEKEK 420
Db 361 IAGGKGINNPSFYLYKEDQLTGSORALSQEBEQNKIDFMEFLAQNNAKLDLSEKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFGNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVFNVDYSNFKYTNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
QY 541 PDNNLAITSFVRNLEDKLTQGLSPQEBANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
Db 541 PDNNLAITSFVRNLEDKLTQGLSPQEBANKLIKDFLSSNNKELVGTINFNKAVADAKN 600
QY 601 TGNVDEVKKAOKLEKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKLEKSLRKHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFALIN 660
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 13:13:01 ; Search time 70.08 Seconds
(without alignments)
1246.750 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 MTNETIDQPPQTEAAFNPOQ.....RENAHGIRKNTKGGCFQKS 1147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5823	99.3	1215	2 B48281	cytotoxin-associat
2	5395.5	92.0	1186	2 C64588	cag pathogenicity
3	5086	86.7	1167	2 B71924	cag island protein
4	349.5	6.0	1302	1 JC6009	surface-located me
5	346.5	5.9	2401	2 T28676	rhostry protein -
6	336.5	5.7	2269	2 T28677	rhostry protein -
7	331	5.6	1127	2 T28317	ORF MSV156 hypothe
8	322.5	5.5	2166	2 G70163	hypothetical prote
9	321	5.5	1365	2 T30822	lmp1 protein - Myc
10	317	5.4	1790	2 S67593	transport protein
11	307	5.2	1252	2 B42771	reticulocyte-bindin
12	302	5.1	2139	2 T18296	myosin heavy chain
13	301	5.1	1957	2 T38077	hypothetical colle
14	294.5	5.0	2523	2 T18477	hypothetical prote
15	293.5	5.0	1005	2 A64465	hypothetical prote
16	291	5.0	1269	2 F84730	probable myosin he
17	290.5	5.0	1738	2 T14867	interaptin - slime
18	290	4.9	1939	2 T18372	repeat organellar
19	290	4.9	1979	2 C71622	hypothetical prote
20	289	4.9	2663	1 S28261	centromere protein
21	286	4.9	1650	2 T18444	hypothetical prote
22	284	4.8	1875	2 S38173	myosin-like protei
23	284	4.8	2285	2 T12796	probable transglyc
24	279	4.8	1679	2 S48385	hypothetical prote
25	277.5	4.7	2288	2 T29999	hypothetical prote
26	276.5	4.7	2829	2 A42771	reticulocyte-bindin
27	276	4.7	1837	2 T41023	probable nuclear p
28	276	4.7	1928	2 S46773	myosin heavy chain
29	276	4.7	4688	2 F82885	hypothetical prote

30	275.5	4.7	2485	1 H71621	serine/threonine-s
31	272.5	4.6	2022	2 T43214	ovtl protein - nem
32	272	4.6	1938	1 MWKWI	myosin heavy chain
33	268	4.6	1208	2 T39068	coiled coil protei
34	267	4.6	1727	2 T50073	myosin-like coiled
35	267	4.6	1805	1 A64224	hypothetical prote
36	265.5	4.5	1413	2 T26467	hypothetical prote
37	265.5	4.5	1940	2 A59287	myosin heavy chain
38	265	4.5	1963	1 MWKW	myosin heavy chain
39	265	4.5	3394	2 T18501	hypothetical prote
40	264.5	4.5	1475	2 T33318	hypothetical prote
41	264.5	4.5	1957	2 A45627	myosin heavy chain
42	264	4.5	624	2 PC6003	surface membrane p
43	264	4.5	1313	2 F96673	hypothetical prote
44	263.5	4.5	3259	1 A56539	giantin - human
45	263	4.5	1164	2 T24806	hypothetical prote

ALIGNMENTS

RESULT 1
B48281
cytotoxin-associated gene A protein cagA - Helicobacter pylori (strains G39G and CCUG
N:Alternate names: immunodominant 128K antigen; immunodominant 135K antigen
C:Species: Helicobacter pylori
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: B48281; A48281
R:Covacci, A.; Censini, S.; Bugnoli, M.; Petracca, R.; Burroni, D.; Macchia, G.; Mass
Proc. Natl. Acad. Sci. U.S.A. 90, 5791-5795, 1993
A:Title: Molecular characterization of the 128-kDa immunodominant antigen of Helicoba
A:Reference number: A48281; MUID:93296225
A:Accession: B48281
A:Molecule type: DNA
A:Residues: 1-1215 <COV1>
A:Cross-references: GB:X70039; NID:g394912; PIDN:CAA49633.1; PID:g394913; GB:X70038;
A:Experimental source: strain G39G
A:Note: sequence extracted from NCBI backbone (NCBIN:133971, NCBIIP:133973)
A:Accession: A48281
A:Molecule type: DNA
A:Residues: 1-957,1026-1215 <COV2>
A:Cross-references: GB:X70039; NID:g394912; PIDN:CAA49633.1; PID:g394913
A:Experimental source: strain CCUG 17874
A:Note: sequence extracted from NCBI backbone (NCBIN:133934, NCBIIP:133968)
C:Comment: These sequences are from cytotoxin producing strains. It is similar to a s
C:Keywords: duplication; tandem repeat
F:958-1025/Region: 33-residue repeats

Query Match	99.3%	Score	5823	DB 2	Length	1215
Best Local Similarity	94.4%	Pred. No.	3.9e-203			
Matches 1147	Conservative	0	Mismatches	0	Indels	68
Gaps	1					
Qy	1	MTNETIDQPPQTEAAFNPOQFINNLOVAFKLVDAVASVDPDQKPIVDKNDKNDNRNQAFEG	60			
Db	1	MTNETIDQPPQTEAAFNPOQFINNLOVAFKLVDAVASVDPDQKPIVDKNDKNDNRNQAFEG	60			
Qy	61	ISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQYRIF	120			
Db	61	ISQLREEYSNKAIKNPTKKNQYFSDFINKSNDLINKNLDIVESSTKSFQKFGDQYRIF	120			
Qy	121	TSWVSHQNDPSKINTRSTRNFNMENIQQPILDDKAEFLKSAKQSFAGIIGNQIRTDQ	180			
Db	121	TSWVSHQNDPSKINTRSTRNFNMENIQQPILDDKAEFLKSAKQSFAGIIGNQIRTDQ	180			
Qy	181	KFMGVFDESLEKROBAEKNGEPTGGDWLDFLSFIDFKKQSDVKEAINQEPVPHVQPD	240			
Db	181	KFMGVFDESLEKROBAEKNGEPTGGDWLDFLSFIDFKKQSDVKEAINQEPVPHVQPD	240			
Qy	241	ATTTTDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS	300			
Db	241	ATTTTDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS	300			

Qy 301 SVLMGSHNGIEPEKVSLLYGGNGGCGARHDMNATVGYKDDQGNVATINVHMKNGLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGGNGGCGARHDMNATVGYKDDQGNVATINVHMKNGLV 360
Qy 361 IAGGEGKGNPVSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLNLSEKEK 420
Db 361 IAGGEGKGNPVSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLNLSEKEK 420
Qy 421 FRTEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYGGKADKA 480
Db 421 FRTEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYGGKADKA 480
Qy 481 LDREKNVTLQGLSKHGDGVNFVDSNFKYTNASKPNKNGVGTNGVSHLEVGNKVAIFNL 540
Db 481 LDREKNVTLQGLSKHGDGVNFVDSNFKYTNASKPNKNGVGTNGVSHLEVGNKVAIFNL 540
Qy 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKAVADAKN 600
Db 541 PDNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKAVADAKN 600
Qy 601 TGNVDEVAKKADLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAOSQKDEIFALIN 660
Db 601 TGNVDEVAKKADLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAOSQKDEIFALIN 660
Qy 661 KEANDARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKAEETLK 720
Db 661 KEANDARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSYKDVIIINQV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSYKDVIIINQV 780
Qy 781 TDVDNLNOAVSVAKATGDFSRVEQALADLNFSKEQLAQQAOKNESLNARKKSEIYQSV 840
Db 781 TDVDNLNOAVSVAKATGDFSRVEQALADLNFSKEQLAQQAOKNESLNARKKSEIYQSV 840
Qy 841 KNGVNGTLVGNLSQAEATTLTKNFSIDIKELNAKLGNFNNNNNGLKNEPIYAKVNRKK 900
Db 841 KNGVNGTLVGNLSQAEATTLTKNFSIDIKELNAKLGNFNNNNNGLKNEPIYAKVNRKK 900
Qy 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVYVGAAGPPLKRHKVDLDSKVGV 957
Db 901 AGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVYVGAAGPPLKRHKVDLDSKVGV 957
Qy 958 ----- 957
Db 961 VSPPIYATIDDLGPPFLKRHDVDDLSKVGRSVSPPIYATIDDLGPPFLKRHDVDD 1020
Qy 958 -----LSRNQELAQIDNLNOAVSEAKAGFFGNLBTQIDKLKSTKHNPMNLWVESAKV 1012
Db 1021 DLSKVLNRNQLAQIDNLNOAVSEAKAGFFGNLBTQIDKLKSTKHNPMNLWVESAKV 1080
Qy 1013 PASLSAKLDNATNSHIRINSIKNGAINKATGMLTOKNPWLKLVNDKIVAHNVGVSVP 1072
Db 1081 PASLSAKLDNATNSHIRINSIKNGAINKATGMLTOKNPWLKLVNDKIVAHNVGVSVP 1140
Qy 1073 LSEYDKIGFNQNMKDYSDSPKFSKLNNAVKTNSGTFQLTNAFSTASYCYCLARENAE 1132
Db 1141 LSEYDKIGFNQNMKDYSDSPKFSKLNNAVKTNSGTFQLTNAFSTASYCYCLARENAE 1200
Qy 1133 HGKKNVNTKGGFQKS 1147
Db 1201 HGKKNVNTKGGFQKS 1215

RESULT 2
C64588
cag pathogenicity island protein cag26 - Helicobacter pylori (strain 26695)
N:Alternate names: cytotoxin-associated gene A protein
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: C64588
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:9739467
A:Accession: C64588
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1186 <TOM>
A:Cross-references: GB:AE000569; GB:AE000511; NID:g2313663; PIDN:AAD07614.1; PID:g231

Query Match 92.0%; Score 5395.5; DB 2; Length 1186;
Best Local Similarity 89.9%; Pred. No. 1e-187;
Matches 1068; Conservative 26; Mismatches 51; Indels 43; Gaps 5;
Qy 1 MTNETIDQ-----QPTAEAFNPQQFINNLQVAFILKVDNAVASYDDQKPIVDKNDNR 55
Db 1 MTNETIDQTRTPDQTQSQTAFDPQOFINNLQVAFILKVDNAVASFDPQKPIVDKNDNR 60
Qy 56 QAFEGISOLREYSNKAIKNPTKKNOYFSDFINKNDLINKDNLIDVESSTKSFQFGDQ 115
Db 61 QAFDGISOLREYSNKAIKNPTKKNOYFSDFDKSNLINKDNLIDVESSTKSFQFGDQ 120
Qy 116 RYRFTSVVSHQNDPSKINTSRINFMENIIOPTDLDKKEAEFLKSAKQSFAGIIGNQ 175
Db 121 RYQFTSVVSHQNDPSKINTSRINFMENIIOPTDLDKKEAEFLKSAKQSFAGIIGNQ 180
Qy 176 IRTDQKFGVGFDESUKERQAEKNGEPTGGDWLDIFLSPFIPDKKQSSDVKEAINQEPVPH 235
Db 181 IRTDQKFGVGFDESUKERQAEKNGEPTGGDWLDIFLSPFIPDKKQSSDVKEAINQEPVPH 240
Qy 236 VQPDITATTTDIOGLPPEAROLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLTH 295
Db 241 VQPDITATTTDIOGLPPEAROLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLTH 300
Qy 296 NNALSSVLGMSHNGTEPEKVSLLYGGNGGPGARHDMNATVGYKDDQGNVATINVHMKN 355
Db 301 NNALSSVLGMSHNGTEPEKVSLLYAGNGGFGDKHDMNATVGYKDDQGNVATINVHMKN 360
Qy 356 GSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLNLSE 415
Db 361 GSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLNLSE 420
Qy 416 KEKEFRTEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYOK 475
Db 421 KEKEFRTEIKDFQKSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYOK 480
Qy 476 KADKALDREKNVTLOGSLKHGDGVNFVDSNFKYTNASKPNKNGVGTNGVSHLEVGNK 535
Db 481 KADKALDREKNVTLOGSLKHGDGVNFVDSNFKYTNASKPNKNGVGTNGVSHLEVGNK 540
Qy 536 AIFNLPDLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKAV 595
Db 541 AIFNLPDLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKAV 600
Qy 596 ADAKNTGNDVKKAKOLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAOSQKDEI 655
Db 601 AEAKTGNDVKKAKOLEKSLRKREHLEKEVEKKLESKSGNKNKMEAKAQAOSQKDEI 660
Qy 656 FALINKANRARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKA 715
Db 661 FALINKANRARAIAQAQNLKGKRELSOKLENVKNLKDFOKDFEPKNGKNKDFSKA 720
Qy 716 EETLKALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSYKDV 775
Db 721 EETLKALGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSYKDV 780
Qy 776 INQKVTDKVDNLNOAVSVAKATGDFSRVEQALADLNFSKEQLAQQAOKNESLNARKKSE 835
Db 781 INQKVTDKVDNLNOAVSVAKATGDFSRVEQALADLNFSKEQLAQQAOKNESLNARKKSE 840

QY 836 IYOSVKNVNGTLYVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 893
Db 841 LYOSVKNVNGTLYVNGLSGIEATLAKNFSDIKKELNEKFN--NNNNNNKLNSTPIY 899
QY 894 ARVKNKKAGQAASLEPIYAQVAKVNAKIDRLNIOASGLGVGQAAGFPPLKRHRDKVDL 953
Db 900 ARVKNKKAGQAASLEPIYAQVAKVNAKIDRLNIOASGLGVGQAAGFPPLKRHRDKVDL 959
QY 954 SKVGL-----SRNOELAQKIDNLNOAVSEAK 979
Db 960 SKVGLSASPEPIYATIDDLGGPFPLKRHRDKVDLSSKVGSRNOELAQKIDNLNOAVSEAK 1019
QY 980 AGFFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGA 1039
Db 1020 AGFFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGA 1079
QY 1040 INEKATGMLTKQNPWLKLVNDKIVAHNVGSVPLSEYDKIGFNQNMKDYSDSFKESTKL 1099
Db 1080 INEKATGMLTKQNPWLKLVNDKIVAHNVGSVPLSEYDKIGFNQNMKDYSDSFKESTKL 1139
QY 1100 NNAVKDTNSGFTQFTLNASTASYCYCLARENAEHGKIKVNTKGGFQKS 1147
Db 1140 NNAVKDTNSGFTQFTLNASTASYCYCLARENAEHGKIKVNTKGGFQKS 1186
RESULT 3
B71924
cag island protein, cytotoxicity associated immunodominant antigen - Helicobacter pylori
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: B71924
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path-
A:Reference number: A71800; MUID:99120557
A:Accession: B71924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1167 <ARN>
A:Cross-references: GB:AE001483; GB:AE001439; NID:94155034; PIDN:AAD06073.1; PID:9415503
A:Experimental source: strain J99
C:Genetics:
A:Gene: cagA

Query Match 86.7%; Score 5086; DB 2; Length 1167;
Best Local Similarity 84.1%; Pred. No. 1.5e-176;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;
QY 1 MTNETIDQOQTEAAFPNQOFTINNLOVAFKVDNVAASYDDPKPIVDKNDNRDNRQAFEG 60
Db 1 MTNEAINOQOQTEAAFPNQOFTINNLOVAFKVDNVAASYDDPKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEVSKRAIKPTKKNQYFDFINKNNDLINKNLDIVESSTKSFQKQDQYRIF 120
Db 61 ISQLEEFANKAIKPTKKNQYFDFINKNNDLINKNLDIVESSTKSFQKQDQYRIF 120
QY 121 TSWVSHQNDPSKINTRSIRNFENIIOPLDDKAEFLKSAKOSFAGIIGNQIRTDQ 180
Db 121 MNWVSHQNDPSKINTOKIRGFENIIOPLDDKAEFLKSAKOSFAGIIGNQIRSDQ 180
QY 181 KPMGVFDESLEKQBAEKNGE-----PTGGDWLIDFLSFDFKQSSDVKEAINQEPVPHV 236
Db 181 KPMGVFDESLEKQBAEKNGEPGPTGGDWLIDFLSFDFKQSSDLKETLNQEPVPHV 240
QY 237 QPDIAATTTDIOGLPEARDLLDERGNSKFTPLGDMEMLDVGVADIDPNYKFNQLLIHN 296
Db 241 QPDVATTTDIOGLPEARDLLDERGNSKFTPLGDMEMLDVGVADIDPNYKFNQLLIHN 300
QY 297 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKQDQGNVATLINVHMKN 356
|||||

Db 301 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDNATVGYKQDQGNVATLINVHMKN 360
QY 357 SGLVIAGGEGKGINNPFYLYKEDQLTQSORALSQBEIQNKIDFMEFLAONNAKLDNLSEK 416
Db 361 SGLVIAGGEGKGINNPFYLYKEDQLTQSORALSQBEIQNKIDFMEFLAONNAKLDNLSEK 420
QY 417 EKEKFTETIKDFOKDSKAYLDALGNDRITAFYSKDKTKHSALITEFGNGDLSYTKDYGKK 476
Db 421 EKEKFTETIKDFOKDSKAYLDALGNDRITAFYSKDKTKHSALITEFGNGDLSYTKDYGKK 480
QY 477 ADKALDRKNTVTLOGSLKHGVMFVDYSNFYKYNASKPNKGVGTNGVSHLEVGFNKA 536
Db 481 ADKALDRKNTVTLOGSLKHGVMFVDYSNFYKYNASKPNKGVGTNGVSHLEVGFNKA 540
QY 537 IFNLPDLNLAITTSVRRNLEDKLTGSLSQEANKLIKDFLSSNNKELVGKTLNFNKA 596
Db 541 VFNLPDLNLAITTSVRRNLEDKLTGSLSQEANKLIKDFLSSNNKELVGKTLNFNKA 600
QY 597 DAKNTGNYDEVKAKQDLKSLRKEHLEKVEKLESKSGNKNKMEAKAQSOKDEIF 656
Db 601 EAKNTGNYDEVKAKQDLKSLRKEHLEKVEKLESKSGNKNKMEAKAQSOKDEIF 660
QY 657 ALINKENRDAIRAIYAQNLKIGIKRELSDKLENVKNLKDOKSFEDEFKNGKNKDFSKAE 716
Db 661 ALINKENRDAIRAIYAQNLKIGIKRELSDKLENVKNLKDOKSFEDEFKNGKNKDFSKAE 720
QY 717 ETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVDVII 776
Db 721 ETLKALKGSKVDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVDVII 780
QY 777 NOKVTDKVDNLNOAVSVAKATGDFSRVQALADLNKFSKEQLAQQAQKNESLNARKKSEI 836
Db 781 NOKVTDKVDNLNOAVSVAKATGDFSGVQALADLNKFSKEQLAQQAQKNEDFTGKNSAL 840
QY 837 QOSVKNVNGTLYVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 896
Db 841 QOSVKNVNGTLYVNGLSQAEATTLTKNFSDDIKKELNAKLGNNNNNNGLKN--EDIY 894
QY 897 NKKAGQAASLEPIYAQVAKVNAKIDRLNIOASGLGVGQAAG----- 941
Db 895 -----TEPIYTVAKKVKAKIDRLDQIASGLGVDGQAASFLLRKHRDKVDLSDSKV 943
QY 942 -----FPLKRHRDKVDLSDSKVLSRQELAQQAQKNEDFTGKNSAL 982
Db 944 GLSANHEPIYATIDDLGGPFPLKRHRDKVDLSDSKVLSRQELAQQAQKNEDFTGKNSAL 1003
QY 983 FGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGAINE 1042
Db 1004 FGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSIKNGAINE 1063
QY 1043 KATGMLTKQNPWLKLVNDKIVAHNVGSVPLSEYDKIGFNQNMKDYSDSFKESTKLNA 1102
Db 1064 KATGMLTKQNPWLKLVNDKIVAHNVGSVPLSEYDKIGFNQNMKDYSDSFKESTKLNA 1123
QY 1103 VKDTNSGFTQFTLNASTASYCYCLARENAEHGKIKVNTKGGFQKS 1147
Db 1124 VKDTNSGFTQFTLNASTASYCYCLARENAEHGKIKVNTKGGFQKS 1167

RESULT 4

JC6009
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JC6009
R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene sys-
A:Reference number: JC6009; MUID:96213016
A:Accession: JC6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336

C;Genetics:

A: Gene: *lmp3*
 A: Generic code: SGC3
 C: Superfamily: surface-located membrane protein *lmp3*; tetratricopeptide repeat homology
 C: Keywords: duplication; membrane protein
 F: 1-24/Domain: signal sequence #status predicted <SIG>
 F: 25-1302/Product: surface-located membrane protein *Lmp3* #status predicted <MAT>
 F: 957-992/Domain: tetratricopeptide repeat homology <TP1>
 F: 993-1026/Domain: tetratricopeptide repeat homology <TP2>
 F: 1089-1120/Domain: tetratricopeptide repeat homology <TP3>
 F: 1154-1190/Domain: tetratricopeptide repeat homology <TP4>

Query Match 6.0%; Score 349.5; DB 1; Length 1302;
Best Local Similarity 19.8%; Pred. No. 1.4e-05;
Matches 243; Conservative 230; Mismatches 478; Indels 277; Gaps 56;

QY	31	KVDNAVASTPDQKPIVDKNDNRNQAEGISQOLREYSNKAIKNTPKNQYFSDPINKS	90
Db	58	KANDFTARQDKKFNSTAFKNHSNTSKLIDBI-----EFSKKILIDQKDEKORLEBNLK	113
QY	91	NDL-INKNDLIDVESSTKSFQKEDQYRIFTSWSHONDPSKINTSRNEM-ENIQP	148
Db	114	NKRLDLONLI-----NSNDGQNVSDSDAKKALNEMQID	148
QY	149	PIILDDKEKA--EFLKSAKOSFAGIIGNOIIRTDQKPMG-VFDESLEKGEAEKNGBPTGG	205
Db	149	SLPIDIKIKTNNENLNAKELL-----NKINABRELQSKIFNE--KKQELRVLDLEDT	200
QY	206	DWLIDFLSFIDFKQSSDYKEAINQ-----EPVPH--VQPDIAITTT-----TD	246
Db	201	KEVD-----FTKEQKVFJETNNTSSIEDTKNKITIEYEKATSSITSLNKTQOEOLE	254
QY	247	IQGLPEAREDLDERGNFSKFTLGDMEMLD-VEGVADIDPNYKFKNOLLTHNNALSSVLWG	305
Db	255	FENIKKLODPFINTKLDNDAKYSIKOKALDKINSLNGINKNSTIKEIKAGQNAL--IKAK	312
QY	306	SHNGIPEKVSLLYGGNGGPGARHDNNAVTGVYKDOOGNNVATTIINVHMKNGSGLVIAGE	365
Db	313	EEAGLEKEK---LDGONIKDITKETINNAKEFKKLLIDNDQKIVD--LKSNDLNEITSKAE	367
QY	366	KGJNNPSFYLKBDLGTGSORALSQOEIOKIDFMFEFLAONNAKLDNLSEKEX--EKFFT	423
Db	368	QSLSKDKESWESANDLLNT-KLIEYKEILNK-----FNOEKEAKFNELEQTRKNLENFUT	421
QY	424	E-----IKDF--QKDSKAYLDALGNDRIAFVSKKD--TRHSALITFEFGNDLSY	468
Db	422	DEVKNNPNVATLVKDLTNAKDAKKSVTNSSN-----KSDIIAANEALIQALDANKA-	473
QY	469	TLKDYCKKADKALDREKNV-----TLOCSLKHGDGMVF--DYSNFKYTNASK--NPNK	517
Db	474	--KQVDEANKSITKEQNLALIDKANTLLPOLNDNDSEIVKAKESLNAEITNANKAVNQND	531
QY	518	GVGVTNGVSHLEVGFNKAIFAIFNLPDLNNLAITSF-----VRNLEDKLTTKGLSPQEAN	572
Db	532	NASMQSAKSSLDKVKYTIQ-NQLTPEKDKDAKFKELEQTRKDIDNFLTDDVKNNPNVAT	590
QY	573	LTKDFLSSNNKELVGKTLPNFNKAVADAKNTGNYDEVYKKAQKLEKSLRKHLE---KEYE	629
Db	591	LVKDLTNAKDDKSVTKSSNKSEIIAAN---DELKQA---LDKAKVAKDQIDEANKSIK	643
QY	630	KKLESKSGHKNKWEAKAQAANSOKDEIFALINK---EANPDARAIYAONL-----KGIG	680
Db	644	EQLSDSITWNLQNLKL-VDSKRD-----IQRAKTELSEIOISASOELNNNPTSMQSAK	697
QY	681	RELSDKLENVNNKLKDFDKSFDEFKNGKNKDFSKABEATLALKLGSVKDLGINPEW---IS	737
Db	698	ESLDKAVTEITKKLETFNKD-----KDVKELEKTRKOIDEIFINTKNTPNVSTLIS	750
QY	738	KYENLNAALNEFNKNGKNRDFSKVTOAKSDLENSVKDVIINQKVTOKV-----DNLNQAVS	792
Db	751	EUTSKRDSKNSVTNSSNK--SDIETANTELQALAKANTDKAQNDAKASTKEQNLNNSI	808

Qy	793	-----VAKATGDFSRVEQALADLNKFSKEQLAQQAQKNSLNARKKSEIYQSVKNGVNGT	847
		: : : : : : : : : : : : : : : : : : :	
Db	809	SANTLLAKLTDKONTIQQA-----KTELEKEVQKANQAVASNNNTASMQSAKSLD--	858
Qy	848	LVGNCLSQAEATTLSKNFSDIKELNKLGNFNNNNNGLKNEPIYAKVNKKKACQAASL	907
		: : : : : : : : : : : : : : : : : :	
Db	859	-----AKVETTKKLETFNKDKDVKPKLEQTR-----KDIDEFINTNK-----	897
Qy	908	EPIYAQVAKVKNAKIDRLNQLIASGLGVVG-QAAGFPPLKR-----HDKVDDLKSL---	955
		: : : : : : : : : : : : : : : : : :	
Db	898	TNPDYSTYLISELSKRDKSNIITSSNKSDIETANTELQALAKAKNTKDQADNLARSTK	957
Qy	956	-----VGLSRNQELAQKIDNUNQAVSBAKAGFGFNLEQTIIDKLKSTKHPNMLWVESAK-	1010
		: : : : : : : : : : : : : : : : : :	
Db	958	EQLNKSISSENTLLAKLTDKDNITIQOAKT-----ELEKEVQKANQAVASN-NTASMQSAKS	1012
Qy	1011	---KVPASLSAKLDNYATNSHIR-----INSNIKNCAINEKATGMLTKGN	1052
		: : : : : : : : : : : : : : : : : :	
Db	1013	SLDAKV-TEITTKLETFNKDKDVKPRELEQTRKDKIDEFINTN-----KTN	1056
Qy	1053	PEWLKLVNDKIVAHNVGSPVLSSEYDVGIGFNQKNKMDKSDSPKFKTLNNAVKDTNSGFTQ	1112
		: : : : : : : : : : : : : : : : : :	
Db	1057	PNYSTLISE-----LTSKRDSKNSIITSSNKSIDIETANTELQALAKAKNTDKAQ	1105
Qy	1113	FLTNAFSTASYCYCLARENAPHGKKNVNT	1140
		: : : : : : : : : : : : : : : : : :	
Db	1106	ADNLARST-----KEQLNKSISANT	1126
RESULT	5		
T28676		rhoptry protein - Plasmodium yoelii (fragment)	
C:Species:		Plasmodium yoelii	
C:Date:		15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000	
C:Accession:		T28676; A45521	
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.			
Mol. Biochem. Parasitol.		76, 329-332, 1996	
A:Title:		Comparison of two members of a multigene family coding for high-mo-	
A:Reference number:		220507; MUID:97077455	
A:Accession:		T28676	
A:Status:		preliminary; translated from GB/EMBL/DBJ	
A:Molecule type:		DNA	
A:Residues:		1-2401 <SIN>	
A:Cross-references:		EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:ABA41263.	
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.			
Mol. Biochem. Parasitol.		42, 241-246, 1990	
A:Title:		Identification of the gene for a Plasmodium yoelii rhopty protein	
A:Reference number:		A45521; MUID:91101660	
A:Accession:		A45521	
A:Status:		preliminary	
A:Molecule type:		DNA	
A:Residues:		2260-2401 <KEE>	
A:Cross-references:		GB:M34281	

Query Match	5.9%	Score 346.5;	DB 2;	Length 2401;
Best Local Similarity	20.0%;	Pred. No. 3.7e-05;		
Matches 303;	Conservative 224;	Mismatches 462;	Indels 525;	Gaps 77;
Qy	21	FINNL-QVAFLKVDNAVASYDPDQKPIVDKNDNRNORAFEGISOLREYSNKAIKPTK- 78		
Db	54	FFNSLGLKLLINKVDS-----DGNIIENDIDN-----FDLSKPKKNFKL 92		
Qy	79	-KNQVFSDFINKSNDLINKDLIDVESSTKSFQKGDQRYBFTSWVSHONDPSKINTRS 137		
Db	93	LEGELNGVFENKNWYKNNKL-----DQFKDTMKLIILLIQYNNFEFGLNDAM 142		
Qy	138	IRNFEMENIQQPLDDKREKAFFLSAKOSFAGLIIGNOITPDQKFMGVFDSGLK-----E 192		
Db	143	TKLKNEGISQKFVYNNQIKQKFDKST-----YDEKKEG-FESLELAKNWE 187		
Qy	193	RQAE-----KNGEPTGGDWLDFL-----SIFDKQSSDVKEAINQ-----E 231		

Db 188 KKKLEIITELKKKNETV---OLDIKIRELIKOIKDIEEQIVNDLKLLELNKKIKEITE 244
QY 232 PVPHV-----OPDIATTTDIOGL---PPEARLDLDERGN-----FSKFTLGDM 273
Db 245 KIEYIKKAVDLKKEKEKONVYIDELAKEPPYQITKYEKKNEITYNTIKSDFDKIYVGDIE 304
QY 274 ML-----DVEGV-----ADIDPNYKEN 290
Db 305 QLYNEMFVQVSEHIEHNKTEITLTKTIDNVYNNIOMMETETVSKLNKINETNNKUS 364
QY 291 QLLIH-----NNALSSVL-----MGSHNGIE---PEKVSU-LYGGNGGPGARHD 330
Db 365 ETILDIKIYIGEITNELNKTLEDKNEKGLSNKIDYAKENVQLNVYKSNILEIKKH- 423
QY 331 WNATVGYKDOOGNNVATIIV-----HMKSGSLVITAGGEGKGINNPSF 373
Db 424 -----YNDQ-----INIDNIKEAKONYDOFKHEMK-----TIPPENMKYQKPSI 464
QY 374 YL-----YKE-----DOLTGS-QRALSOEI---ONKI--- 397
Db 465 EIKIMKDEFLSKVKNYNDPDKYKKEVSEHNKFTLTNWKITEVSEDEIKKYEKNFND 524
QY 398 -----DMFEFLAQNNNAKLDNLSEKE---KEKFRTEIKDPQ 429
Db 525 KSLINETKKSIBEEYONINTLKKVDDYIKVCLNTNELITNCHNQTTLADKLNQNKTIK 584
QY 430 KDS---KAYLDALGDNDRIAFVSKDKTKHSALIT---EFGNGD-LSYTLKDYKGGKADKALD 482
Db 585 ETNSIDKIYTDKFENILTDKKTELETFTKGLSLNNHESNNKELLTYF---YDLKANLGN 641
QY 483 REKNYTLQOSLKHGDMFVDYSNFKYNASKNPNGKGVGTNGVSHLEVGFKNVAIFNLP- 541
Db 642 KENMLYKOPNEKEKAVEDI-----KKKNVDINKIVSNIEITIT-YTSYININE 687
QY 542 -----DLNLAITSFVRN-----LEBKL-----TTKGLSPOEANKLIK 576
Db 688 DTENEIGKSIELNLTVKLEKVANVTNLEIKEKLDYDFDFGKEKNKYDPENKIKND 747
QY 577 FLSSKNELVGKTLNFKNAVADAKN-----TGNVDEVKKA-----QKDLKSLR- 619
Db 748 IDTLNOKIDKSIETLTETKIKKSENHIDEIKQIDKLKVPNKTMTFNEDEPKETEKENIV 807
QY 620 ----KREHLEKEVEK-----KLES-----KSGNKKMEAKAQA- ---SQDEIFALIN 660
Db 808 EKIDKKNKIYKEIDKLLNLEISINDKTSLEKLNKLNLSYKSLGNLFLOQIDE-----EK 863
QY 661 KEANRDARAI-AYAQNKLKIKRELSKLE-----NVNKNLKDFDKSGFDFEPK--- 705
Db 864 KKAETHIKAMEAYIDDLONIKKK-SOEIEKEMNINMDIKMDIHKENKALNISHDDYKIIYH 922
QY 706 -NGKN-----KDFS-----KAEETLKALGSKYKOLGINPEWISKYE 740
Db 923 TTSKNHEEKISDIRKNSLIIQDFSESVINDIKKEKLVLESQNNNTDIN-QYLSKIE 981
QY 741 NLN--AALNEFNKNGKDFSKYTOAKSDLENSVKVVIINQKVDKV-----DNLNOAV 791
Db 982 NYNITLKNKIK---KIIDKVEKYTDETEKNKKKINAELSSEKIITQKENSLSKECQ 1037
QY 792 SVAKATGDFSRYEQALADLKNF-----SKEOL-----AQQAQKNESLN-----ARKK 833
Db 1038 SKIKSTIDNYYSECIKNITNLTUKTYIVNEKNINITYFKNAEEVNVNLSFNFNIEADTK 1097
QY 834 SEIYOSVK--NGVNGTFLVGNLSQAEATTLSKNFSDIKKELNAKLGNNFNNNNNKLNKP 891
Db 1098 SQVILNKKNGTNTDYN-----IKELHEKKKSNVYKDEACKNTQEIKNKE 1146
QY 892 IYAVKN-----KKKAGQAAASLEPIYAQV-----AKVN--- 920
Db 1147 LFEKVEQOEVTLLNKYIAVELNKKFKDKTKNYSQEIILKEIKDAHNTFTSQADKSEKKMNEI 1206
QY 921 -----AKIDRLNQIASGLGVGQAAGFPLK-RHDKYDDL---SKVGLSRNQELA 965
Db 1207 KNEQIRIEDEAVANNKSNKAILDI-----QLSVEPPKIKFLKTKDITRKSDDCCLKETKDIE 1262

QY 966 QKIDNUNQAVSEKAGFFCNLEQFIDKLDKSTKHNPMLNWSAKKVPASLSAKLDNYAT 1025
Db 1263 TKISNLSIDTQETKLTENKNILNLTLEKLSLKNOKN--IEDOKKELDEVNSKIKNIES 1320
QY 1026 NSHIRINSKINCAINEKATGMLTOKNPWLKLVDN-----KIVAHNNGSV--PLSEYD 1077
Db 1321 N-----NQHKKNVEIG-----IVEKINEIAKANKOIESTQKLIITPTKNLSPPKAND 1370
QY 1078 KIGFN-QKNMKDYSDSFKEFTKLNNNAVKOTNSGTOFLTNAFSTASYICLARENAEH--- 1133
Db 1371 LEGIDTNKNLG-----KYNTENMNIYEEFIKSY-DLITHYLETVSKEPITYEQIKNKRI 1423
QY 1134 -----GIKNVN 1139
Db 1424 TAQNELLTNIKVN 1437
RESULT 6
T28677
rhoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
R:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii
A:Reference number: Z20508; MUID:95021522
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KE>
A:Cross-references: EMBL:L27838; NID:g457145; PID:g457146; PIDN:AAA21304.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A:Reference number: A45521; MUID:91101660
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match 5.7%; Score 336.5; DB 2; Length 2269;
Best Local Similarity 19.5%; Pred. No. 8e-05;
Matches 243; Conservative 232; Mismatches 479; Indels 291; Gaps 59;

QY 22 INNIQVAFILKVDNAVASYDDQKPIVDKNDNRQAFEGISQLRREYSNKAINKPTKKNQ 81
Db 234 LNNIETN-NKLPTILEI---KRYIYDEISKELNKLMLDFKKEKELSNK-ISDYDKRE 288
QY 82 YFSDFTKS-----NDLINKDLIDVESSTKSFQKFGDORYRIFTSWYSHONDPSKI- 133
Db 289 QLSEYKSKMLEIRNHYNSTQNTVDNTEK-EKAKONYDKSNEH-----MTTPTNDEDESKII 343
QY 134 -NTRSRNFMENIQQPIILDKEKAEFLSAKOSFAGIIIGNOIRTDQKPMGVFDESKE 192
Db 344 SEVKTMKDELTSKVNTYIDFNKKYKETVNSEHSQFTEL-----TDKIKAEVSDRELKK 396
QY 193 ROEAENGPEPTGGDWLIDFLSFIDFKQSSDVKEAINOEPVHVQPDIAATTDDIOGLPP 252
Db 397 CEQ-----SFNDKNSLNETKNSIEK-----YONINT-----LKKVDE 430
QY 253 EARDLLDERGNFSKFT---LGDMEMLDVEGVAD---IDPNY--KFNOLLIH---NN 297
Db 431 YIKVCKSTKESITKFSKQTKILKMLNQNIKTYKETSNDKSTIERKPEQLITGQTKLEN 490
QY 298 ALSSVLGMSHNGTEPKVSLLYGGNGGPGARHDNATVGY-----KDOQGNV 345
Db 491 KTEFSLNNHEANNNELIKYF-----SDKANLGINEENMLYNQFTEKEKTFNDI 540
QY 346 ATIIIVHMKNGSLVIAGGEGKGINNPSFYLYKEDQITGSORALSQBEIONKIDFMEFLAQ 405

Db 541 KE-KNIHINE---EISKIEIKI-HASIYNISEE--TEREIGINIESLNTKV--FEKVKR 590
Qy 406 NNAKLDNLSEKEKEKFRTEIKPOFODSK--AYLDALG--NDRIAFAVSKDKTKHSAITEPG 462
Db 591 NVTNLNKEKEKLK---HYDFSFGEGNTKYTDKIKINDIMAVSQQIDQHINGLDDIQ 647
Qy 463 NGDLSY--LKDYGGKADKALDRE-KNVTLOGSLKHGVMFYDYSNFKYTNAKNPNKGV 519
Db 648 KKSSEYVSEMKEQINKLEKVSNTSEISDNVVEGKKKKQIIVTKID--KKKNIYEINKLL 705
Qy 520 GVTNGVSHLEVGNKVAIFNPLDNLNLAITSFVRNLEDKLTGKLSPOEANKLTKDFLS 579
Db 706 SEISKIEKDNSTLEKVKDNLNGYQNL-----GNLFLEQID 741
Qy 580 SNKELVGKTLNFKAVADAKNTGNVDEYKAKOKLEKSLRKRHELEKEVE-----K 630
Db 742 EKKKAENTIKWEAVID-----DLONIKKQSEIETEDIMDKMDINKEMEALKISHDDQ 796
Qy 631 KLESKSGN-----KNKMEAKAQAANSOKOEIFALINKE 662
Db 797 KCHDKSKNHKENISDIYDKSSKIIQDFSRESINDIKNKLQ-KNYESQNH-----NSD 849
Qy 663 ANRDARATAYAO NLGKIRKRELSOKLENVNKLUKDFDKSPDEFKNGKNDKDFSKAETLKA 722
Db 850 INOCLNEVANIYILK-----NKIKTIIDKYKEYTSEIEKKNKINDELNLSERKVIK 904
Qy 723 KG--SVKDL--GINPEWISK-----VENLNAALNEFKNGKNDKFSKVTOAKSDLENSVK 773
Db 905 EGDLSLKECRKINSTLDDKIDECIKINVLK---KNILNEE-TNITNHFNAEYNI 960
Qy 774 VIINQKVDKVDNLQAVSAKATGDFSVEQALADLKNFSKEQLAQQAQKNESNA-RK 832
Db 961 VLSNFNNIEMADNKQYILEIKKNGTNDHDYNIKELSHKDKSNGYKTEADQONKKAIOK 1020
Qy 833 KSEIYQSVKNGYNTLVGNLSQAEATLTSKNFSDIKELNAKLGNNFNNNN-----884
Db 1021 NKELFEQYKEEV--TVL---LNKYAVELNKNFKDKTKNDKSDKQIIEIKDAHNYCTLES 1075
Qy 885 -----NLKNEPIYAK---VNNKKAGQA--ASLEEPIYQVAKKVNKADRLNLOIASGLV 935
Db 1076 SEKKNEIKNEKIHIEDEVANNKSNKAITSIKSV-----EPFKTKIKINEIT---- 1126
Qy 936 VQQAAGFLPKRHDKVDLDSKVLGSLNQELAQKIDMLNQAQVSEAKAGFGNLEQTIDKLKD 995
Db 1127 -----KSDCC-----LKETNDLEKQISLSLTDQETKLTENGKQLTLEELLE 1169
Qy 996 STKHPMNLWVESAKKVPASLSAKLDNYATNSHIRNSIKN-----GAINEKATGMLT 1049
Db 1170 SUKKOKKN--IEDQKKEDELVDNSKIKNTENT-----VNOHKKNYEIGIVEKINEIAK---T 1220
Qy 1050 QKNP-EWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMK--DYSDSF-KFSTKLNNAVKD 1105
Db 1221 NKNOIESTKELIKPTIOHIISS-----FNANDLEGIDSDENLKYNTFMGNIYEE 1270
Qy 1106 TNSGTQELTNFASTASYCLARENAEH-----GIKNVN 1139
Db 1271 FTKSY-NLITNVLETVSRESITYIQNKRIDTQKELLKNIEVN 1314

RESULT 7
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 5.6%; Score 331; DB 2; Length 1127;
Best Local Similarity 20.7%; Pred. No. 5.4e-05;
Matches 249; Conservative 177; Mismatches 440; Indels 338; Gaps 53;

Qy 53 DNRQAFEGISQUREEYSKAIKNPTKKQYFSDFINKSNLDINKONLIDVESTSKSFQRP 112
Db 25 NNVSLDIINSLEYILNN--IK-----FSDKI--TNEIKKNYKI-VERIFYMHQF 71
Qy 113 GQORYRFTSM-VSHONDPSKI--NTRSIKRNFMENIIQPPIL-----DOKERAEFLK 161
Db 72 KINDYNILOIYDIEYNEINEIKCIENKPKCKNPLNYITTKKLYIYDLDYEEKKOKELVI 131
Qy 162 SAKQSPAGIIG-----NQIRTDQKFMGVFDESLEKQERQAEKNGEPTGDMLDIFLSFI 215
Db 132 NIEQKNAVDKINDIKNNVNNIHSNDETIITGKETL-----IDILNKL 174
Qy 216 FDKKSSDVKEAINQ-----EPVPHVQDPDIATTTDIOGLPPEA-RDLDERG 262
Db 175 --KLVSDEKQLEIQIYKNNKNEKEIEFKNIDNVQREINKQDELNKLDESKEFTKQ 232
Qy 263 NFSKFTLGDMEML-----DVEGVADIDPNYK-FNOLLIHNNALSSVLMGSHNGIEPKVS 316
Db 233 ELNKYIDKKQEBELIKKLNKDEKFNFNIDEKQKLLDQINSKINTLNENIKVMNLYETK-- 290
Qy 317 LLYGNGGPGARHWDNATVGYKDOQGNVYATIIVHMKNGSLGVIAGGEGKINN-PSFYL 375
Db 291 -----NKISNLQNEILNKDSTIKSLDEKQ---KLLDELKNNITSLYN 332
Qy 376 YKEDLTQSQRALSQEEIQNKIDPMEFLAQNNAKLDNLSE-KEKEK-FTEIK-----DQ 429
Db 333 KSNITAITIQOLLE-----SSLTDF-----NNANI-NINELSKIKLEFNDIOKLNNDIT 381
Qy 430 KDSKAYDALGNDRIAFYKSKDKTHSALITEFGNGDLSYTLKDYGKADKADOREKNVTL 489
Db 382 EQNNKITDFFNNSTRIFREKLDTEYKK-IDDIKNNNLQLEESY-KKIDQETEYKNNKI- 438
Qy 490 QGSLKHGDMVFDVSNFYKTYTNASKNPNGVGTNGVSHLEVGNFY--AIFNLPDLNLLA 547
Db 439 -----NKEYNDIIEELNKNLQKLEENKIDQETEYKNNKINKEYNDIIEELKN-NNLQ 490
Qy 548 ITSFVRRNLEDKLTGKLSPOEANKLIDFLSSNKELVCKTLNF-----NKAVADAKNTGN 603
Db 491 KLEENKNINDKLTK-----LNKDIESNTSELPNK-LNISDFKDKSREIAKLNTE 538
Qy 604 YDEVKKAOKLEKSLRKRHELEKEVEKKLES-----KSGNKNKWE-----AKAANSOK 652
Db 539 YEOLR---KDLLENINKTNELMKLSDNKLSLEQLYDSKKNILDGIDIKTYNSLSEKNDKI 595
Qy 653 DEIF-----ALINKEANRARAIAAYAQNLKIGIKRELSKLENNVN 691
Db 596 DEYFSNIEKFDIYNNVIENKFNIGNLDLSIINKIINDDQ-----PREYINSKIDS 644
Qy 692 KNLKDFDKSDFEFKNGKNDKFSKAEETLKALGSKVKGKLG-----INPEWISK----- 738
Db 645 NELSTM---FDDIFNAKQ--IASITNNIENSNKIKDLNEFIISNEDSSKLELDEIRKYK 700
Qy 739 -----VENLNAALNEFKNGKNDKFSKVTQAKSDLENSVKDVIINQKVTDKVDNLNQA 792
Db 701 QQFDKIKDAMNTEVSKFENTLOKDIDSISKNINELTNAYD--IINTKANDLDDKLNNGYS 758
Qy 793 VAKATGDFSVEQALADLKNF--SKEQLAQQAQKNESLNARKKSEIYQSVKNGVNTLVG 850
Db 759 -----EFKNLYNNASDLDTIQKNDEKVKQOLNEYLEKNNK----- 794
Qy 851 NCLSQAEATLTSKNFSDIKELNAKLGNNFNNNNGLNNEPIYAKVNNKKAQAASLEBP 910
Db 795 -----QSIEINDIVNNFIKELIKFNNTETNKSLELL----- 826

Qy 911 IYQAVKVNKIDRLNQLASGLGVGQAAGFPLKRRHKVDLDSKVLGRNQLAQKIDN 970
Db 827 ---TNDNDIKIFKL-----YKELNKISTNNLLKTYRNE-----IDN 860
Qy 971 LQAVSEAKAGFGNLEQTIDKLKSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIR 1030
Db 861 VNEKLS-----IVENLQFINSLFSEFFNOGTSITSHN 893
Qy 1031 INSIKNG-----AINEKATGMITQKNPEWLK-LVNDKIIVAHNVGVSPLSEYDKIGFNQ 1083
Db 894 FLLNTLAGINDVNLKLNKIMADTTRGDTNIRDEIKNOISSENKISQK-----PNE 945
Qy 1084 KMKDYSDFKSTFKLNNAVDTSGTQFLTNAFSTASYCLARENAE-----HGI 1135
Db 946 KNEKDLKLISFNDKLNK--YNISAGYTEYNNIEHCLKLYLIAVSDQERYRYFIHNI 1003
Qy 1136 KNVN 1139
Db 1004 KQID 1007

RESULT 8
G70163
hypothetical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PIDN:AAC66876.1; PID:g268842
A:Experimental source: strain B31

Query Match 5.5%; Score 322.5; DB 2; Length 2166;
Best Local Similarity 19.1%; Pred. No. 0.00024;
Matches 259; Conservative 246; Mismatches 496; Indels 357; Gaps 63;

Qy 6 IDQOPQTEAENPQ--QFINNLQVAFKVDNA-----VAS---YDPQKPIVDKND 51
Db 129 VDKLSKTLKGFMTQIDSVESNLNVLEKFKDKANKENLESIKASWEKFDTNIKELVFKID 188
Qy 52 RDNROAF---EGISOLREYSNKAIKNPTKKNQVFSDFINKSNDLNKIDNLDIVESSPKS 108
Db 189 NLNKEISLYKDLANIERKKNDILVKGNEKLDLDFESDFLEKVEFNIGYSK-EIESSF-- 245
Qy 109 FOKFGDORYIFTSWVS--HQNDPSKINR-----SIRNFMENIIPPLDDEK 156
Db 246 --NVENYKLIENSIELIMESVKNKINEKEDFLNRLNEELQNKFKDIL--VYVDDRSK 301
Qy 157 A-----EFLSAKQSFAGIIG--NQIRTDQKF-MGVFDES-----KERQE 195
Db 302 EIKDKLEKLVLDVNETSSSSSKDNVYSRINSLESMEKRYEQVDVDFDKFRSQ 361
Qy 196 AEKNEPTGGDWLDFLFIFFDKQSSDVKEAINOEPVPHVOPDIATTTTIDQGLPPPEAR 255
Db 362 VELNLKNIYEDYED-----KISOVDNIRE-----RVLSLLDLSNMESVQSACI 407
Qy 256 DL---LDERGN-----FSKFTGLDMEMLDVE--GVADIDPN-----YK 288
Db 408 DFKRLEDDSGIVLEFKGFKGADIEVFSFKGIDNQLKQWLESQDLVDVDSNIQEKLIK 467
Qy 289 FNQLLIHN-----NALSIVLMGS-----HNGIEP----- 312

Db 468 LNDNLISNFEIEINGRNNYNSLNDNINAKYATALPESLDSSSKFENQMESKYKSFTDKL 527
Qy 313 ---EKVSLLYGGNGGPGCARHDWNTATVGYKDOQG-----NNVATIIIVHKN----- 355
Db 528 TAGMDEFSLMTGKFKETLSQ---EATNNYQFQDLNKLKLENEIESFYNNFEKTQETLKVD 584
Qy 356 -GSLVITAGGEGKGNPSPFYLYKEDOLTSQORALSQOEIQNKIDPFMEFLAQNNAKLDNLS 414
Db 585 FNTSLINTKDEIGNIVEF---RDRIYDEWNIFFVQLEESKLQYKQWQEMDSNLKNI 640
Qy 415 ---EKEKEKFEIKDFQKSKAYL-DALGNDRIAFVSKK-----DTFKHSAL 457
Db 641 SOIKNTNEEFLSLIQ-IQDKGIELSESVFNDSDHIQKALDMHGSWKDELIALNKSIL 699
Qy 458 ITEFGNDL--SYTLKDYKKADKALDKREKNVTLQ-GSLKHDGVNMFVDSYKFNKYTNASKN 514
Db 700 DIKVSSELLSSATLKIESLEK-VNDRMEYVLLKTGDI--ESLVIEKYKELKDMYSQS 756
Qy 515 PNKGVTNGVSHLEVGNKVAIFNLPLDNLNLAITSFVRRNLEDKLTTKGLSPQEAANKLI 574
Db 757 DEAILGIKEFINRQTEIHKDSVFMLEDLN-----KXFDKKNFVSKIEEDCYKL 807
Qy 575 KDFLSSNKELVGKTLNFN-----KAVADAKNTGNY---DEVKKAQDKLESLRK 620
Db 808 KDFKIESEDILN---NFKSDLNFEIESKLQIVSNIKSNQKQIDFLDRISKDI---LNR 861
Qy 621 REHLEKEVEKKESKSGNKNKMEAKQ--ANSQKDEIFALINKEANRARAIAAYA-ONLK 677
Db 862 KDSINNEVDSLKSLDWOSKQNEITVKIENLLSSGKVDL-DLIDSEVTTKIKELKFSIESLE 920
Qy 678 GIKRELSDKLENVKNLKDPKSDEFKNGKNGKDFSKAEETLKALKGSVKDLGINPENIS 737
Db 921 SYYLEKIDEFNRQAGIYS--DELLQDINMHFNKETRELEENLSKKFAAV--LNNSEEVK 976
Qy 738 KV-----ENLNAALNEFKNGKNGKDFSKV-----TQAKSDL 767
Db 977 EVDLSLQDKRTDIASFOANIDITLDSLVNKFNDINKEINGKYNEVISNRYGSENISSKL 1036
Qy 768 ENSVKDVI--INQVTKDKVDNLQA---VSVAKATGDFSRVEQALADLNKFSKQLAQ 821
Db 1037 ENEMHETENLSRRLTDRIKSLSGMDENLOKLKESFDVSKYQVEKFEK-----VKDL 1090
Qy 822 AOKNESLARKKSEIYQSVKNGVNGTL-----VGNLSQAEATTLKSNFSDIKKELNAKL 876
Db 1091 TDDGEAKTNKLVKETEYQYKSRLEADIDYRRITDINDIMQAK-----ERFGEITNELKNNI 1145
Qy 877 GNFNNNNGLKN-----EPIYAKVKKKQAGASLEPIYAQVA---KKVNAKID 924
Db 1146 ESKSEFLNDLYKERFKLIESNFEEYRSTFLIESEGAISKIRDEIYKTLTSDNENLQIKIS 1205
Qy 925 RLNQIASGLGVVGOAAGFPLKRRHKVDLDSKVLGRNQLAQKIDN---LNOAVSEAKA 980
Db 1206 EMDQ-----NFEILEQSRKIDLEFEKELQDKIKDCYCFINSQFGEIKA 1248
Qy 981 GFGNLEQTIIDKLKDSKTHNPMNLWVESAKKVPASLSAKLDNYATNSHIRNS----- 1033
Db 1249 GVEENIKNHFD-----VCIKKVTNLDIDIVKYENEIHKRIDSLKSIEST 1293
Qy 1034 --NIKNGAINKATCML---TQKNPEWLK-----VNDKIVAHNVGVSPL 1073
Db 1294 FDSIEKN-LNDKVSQCIDKIANDFNLKYLEEERNEGOLNLENKIDNKIKR--IDNLAL 1350
Qy 1074 SEYDKIGFNQNMKDYSDSF-KFSTKLNNAVKTNSGF 1110
Db 1351 SOYDGL-----EKYADMYDEFSERLNSYIATLSEEF 1382

RESULT 9
T30822
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822

R.Jensen, L.T.: Ladefoged, S.; Birkelund, S.; Christiansen, G.

Infect. Immun. 63, 3336-3347, 1995

A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the pre

A:Reference number: 218884; MUID:95369882

A:Accession: T30822

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1365 <JEN>

A:Cross-references: EMBL:U21962; NID:g790243; PID:g790244; PIDN:AAA81013.1

C:Genetics:

A:Gene: lmp1

A:Genetic code: SGC3

Query Match 5.5%; Score 321; DB 2; Length 1365;

Best Local Similarity 22.3%; Pred. No. 0.00016;

Matches 244; Conservative 165; Mismatches 413; Indels 274; Gaps 58;

QY 58 FEGISOLREYSNKAIKNPTTKNOYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDORY 117

DB 124 FNEIKDKLOEY----IKNELSKOEYEHKQNTENEL-NKYTPISLESTLIEIQATNNLI 178

QY 118 RIFTSVSHONDPSKINTSRNEMENIOP----PILDDKEKAFLKSAKQSFAGIIG 173

DB 179 KLLNESTREKONIDNLNAK--EQLKASISQANQLLPQSLDND-SEIAKAKSLDAEIKNA 235

QY 174 NOIRTDQPMGVDFDESLEKROBAEKNGEPTGDMWIDFLSFYDFKK-QSSDVKEAINEQEP 232

DB 236 NO-----AVASNNTASMQSAKSLDAKVAEITKLETENKDKKAKFNLKQTRNQ-- 285

QY 233 VPHVQPDIAITTTDIOGLPPEARDLIDE-----RGNFSFTGLDMEMLDVEGVADID-PNY 287

DB 286 ---IQEFTINTKNN-----PNYSELISQTSKRDSKNSVT-----DSSNKSDESANT 330

QY 288 KFNQLLIHNLSSVLMSHGNGIEPEKVSLLYGGNGGPGCARHDMNATVGYKDDQGNVAT 347

DB 331 ELKQALAKANA-----DKVOA-----DMLAKSIKEQLNNSVSN 363

QY 348 IINVMHNGSLVGIAGGEGGINPFSYLYKE-----DQLTGSORALSQBEIQNKIDFMBFL 403

DB 364 ANTLAK-----LTDKNTIQAKTELEKEVQKADQAIKSNNTASMQSAKSLDAK--V 415

QY 404 AQNNAKLONLSKEKEF-----RTEIKDF---OKDSKAYLDALGNDRIFAVSKKDTKH 454

DB 416 ABITTKLETENKDKKAKFNLKQTRNQIEFTINTKNNPNYSELISQ-----LTSKRDSKN 471

QY 455 SALITEFGN-GDLSYTLKDYGGKADKA-LDREKNVTLOGSLKHGDMFVDYSNFKYTNAS 512

DB 472 S--VTDSSNKSDESANTELKQALAKANADKQVADNLAKSIKEQLNNSVSNAN---TUSA 526

QY 513 KPNKGVGTNGVSHLEVGFNKKV--AIFNLPDLNLAITSFVRNLEDKLT--TKGLSPQ 568

DB 527 KLTDKNTIQAKTELEKEVQKADQAIKS-----NNTASMQSAKSLDAKVAEITKLETF 582

QY 569 EANKLIK-DFLSSNKELVGTNLNFKAVADAKNTGNYDEVKKAQDKLEKSRKREHLEKE 627

DB 583 NKDKAKEFNLKQTRNQIEFTINTN-----KNPNYSEL-----ISQTSKRD----- 625

QY 628 VEKLESKGNKKEAKAQAQNSOKDEIFALINKENARDAIAVAQNL-KGIKRELSDK 686

DB 626 -SKNSVTDSSNKSDESANT-ANTELKQALA-----KANADK---VOADNLAKSIKEQLNNS 674

QY 687 LENVN---KNLKDFDKSFDEKNGKNDKDFSKAEETLK-----ALKGSVKDLGIN-PEWIS 737

DB 675 VSNANTLSAKLTDKNTIQAKTELEKEIQKANOAIKSNNTASMQSAKSLDAKVAEITK 734

QY 738 KVENLN-----AALNEFK-----NGKNKDFSKVTO---AKSDLENSVKDVIINQ 778

DB 735 KLETFNKDKKAKFNLKQTRNQIEFTINTKNNPNYSELISQTSKRDSKNSVTDSSNKS 794

QY 779 KVTDKVDNLNCAVSAKATGFSVEQALADLKNFSKQEAQAQAKNESLNARKKSEIYQ 838

DB 795 DIESANTELKQALNTAKAK--SSIDNELRPLKN-----DLQSKIIEFG 836

QY 839 SVKNGVNGTFLVGNLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNNGKLNPYAKVNK 898

DB 837 PIRN-TNFSWISSKLE---TTKNLAELTKA-----DAIKNNPSSSKQAL 878

QY 899 KKAGQAASLEEPYIAQVAKKVNAKTDLRLNQIASGLGVGQAAAGFPLKRHDKVDLDSKVGL 958

DB 879 KDSSQ-----QVOKLGN---ELLKTIIEEFG-----KVE----- 904

QY 959 SRNQELAKIDNLNCAVSEAKAGFEGNLEBQTIDKLKDSKTHKPNMLNWSAKKVPASLSA 1018

DB 905 TKNISGIRFLFLAQAEQ-----FNNSD--VDKLUK-----KOTLLSKKQ 946

QY 1019 KLDNYATNSHIRINSNIKNGAINERKATGMLTQKNPEWLKLVNDKIVAH---NVGSPVLS 1075

DB 947 KLGNSTKDY-----LTOLSTEMSTQEST--IKKVIVNIQAHIRNLSQYRLE 993

QY 1076 YDKIGFNOKMKDYSD 1091

DB 994 ADKLIANKM--RGYGD 1007

RESULT 10

S67593

transport protein USO1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D2552; protein YDL058w

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000

C:Accession: S67593; A38455; S30782

R:Blöcker, H.; Brandt, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67587

A:Accession: S67593

A:Molecule type: DNA

A:Residues: 1-1790 <BLO>

A:Cross-references: EMBL:Z74106; NID:gl431058; PID:e253003; PID:gl431059; MIPS:YDL058

A:Experimental source: strain S288C

R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.

J. Cell Biol. 113, 245-260, 1991

A:Title: A cytoskeleton-related gene, USO1, is required for intracellular protein tra

A:Reference number: A38455; MUID:91185402

A:Accession: A38455

A:Molecule type: DNA

A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>

A:Cross-references: GB:X54378; NID:g4777; PIDN:CAA38253.1; PID:g4778

A:Note: the authors translated the codon ACT for residue 768 as T

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.

submitted to the EMBL Data Library, February 1993

A:Description: An integrin analogue in Saccharomyces cerevisiae.

A:Reference number: S30782

A:Accession: S30782

A:Molecule type: DNA

A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580

A:Cross-references: SGD:S0002216; MIPS:YDL058w

C:Genetics:

A:Gene: SGD:USO1; INT1

A:Cross-references: SGD:S0002216; MIPS:YDL058w

A:Map position: 4L

C:Keywords: coiled coil; transmembrane protein

F:326-342/Domain: transmembrane #status predicted <TM1>

F:394-410/Domain: transmembrane #status predicted <TM2>

F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 5.4%; Score 317; DB 2; Length 1790;

Best Local Similarity 18.5%; Pred. No. 0.0003;

Matches 233; Conservative 213; Mismatches 467; Indels 344; Gaps 48;

QY 3 NETIDQQPOTAEAFNPQOFINNQLQVAFKVDN-----AVASVDPDOKPIVDKNDNR 55

DB 677 NEDSILPELDETGLPKVIFSTY---FIQLFNENIYRTRTALSHDPDEEPI-----NK 726

QY 56 QAFEGISOLR-----EYSNKN--AIKNPTKKNQYFSDFINKSND 92


```
Db 508 ITNIEGALKESKONYEIGF-----LEKLEETG-----KNRKLKVDITKCSI-----NST 551
QY 574 IKDFLSSNRELKVTILNFNKAVADAKNTGNYDEVKKAQDLEKSLRK-REHLEKEVEKKL 632
Db 552 VGNFSSLENNFDLNOYDFNKNINDYEN-----KMGIIYNEFEGLSNKISENLNASENTS 606
QY 633 ESKSGNKNMEAKQAQNSOKDEIFALINK--EANDRARAIAQAQNLKGTGKRELSKLENV 690
Db 607 DYNASAKTLREA-----OKERV-NLLNKEEANKYLROVKKVESPREI-FNNKESLDDKI 658
QY 691 NNKLDKDFKSFDEFKNGKNKDSKAEETLKALKGSKVKGDLGINPEWLSKVENLNAALNEFK 750
Db 659 NEMIK-----KEQITVNEGHVGNVQJVENITKELVDENNLSDLKQ-A 699
QY 751 NGKNKDFSKVTOAKSLENSVKDVI-----INQKVTD----- 782
Db 700 TCKNEISQIKTH--STLKNKAKTILGHVDTSAKYVICIKITPELALTELLCDAKLTAQEL 757
QY 783 KVDNLNOAV-----SVAKATGDFSRVBQALADLNKFSKEOLAQO-----AQKNESLNARKK 833
Db 758 KPESKNNVLETEMNSKNTNELD-VHKNIQDAYKVALEILAHSEIDTDTKODKSSKLIEMG 816
QY 834 SEIYQSV-----KNCVNGTLVGNLSQAQEAATLTKNFSDIKK-----ELN 873
Db 817 NQIYLKVLINQYKKNKISSIKSKEEAUVKIGNVSKKHSLSKITCSDKSYDNIILALEKQ 876
QY 874 AKLGNFNN-----NNNN-----GLKN--EPIYAKVNNKKAG-----QAASL 907
Db 877 TELONLRNFTQKWTNWSDSKLEIKITDFESLKNALKTLEGEVNAKASSONHEHVQSK 936
QY 908 EPIYAQVA--KKVNAKIDRLNQIASGLGVGQAAGFPLKRHRDKVDLDSKVGLSRNQELA 965
Db 937 SEPVNPALSEIEKEETDIDSLTALDELKKGRTG--EVSRYKLKIDVTYKELSDDEL- 993
QY 966 QKIDNLNOAVSAAKAGFFNLEQTDKLT-----KOSTHNPMLNWSAKKVPASL 1016
Db 994 --INTIEKVVKAYLAYIKKNYEDTVQDVLTLEHFNFTKQVSNHEPTNF--DRSKN-----S 1045
QY 1017 SAKLDNYATNSH-----IRNSNIKNKAINEKATGMLTKQNPWLKLVNDKIVAH 1066
Db 1046 SEELKAVTDSKTIISKLGVIIEVENTEMTITIESSAKEIEALYN-----ELKNKTSJN 1101
QY 1067 NV-----GSVPLSEYDKIGFNQKNMDYSDSPKFTSKLNNNAVDTNSGFTQFLTNAPFSTAS 1122
Db 1102 EIYQTSNEVKLOE-----MKSADKYIDVSKIFNTVLDTKQ--SNIVTN----- 1143
QY 1123 YYCLARENAEHGKKNVNTK 1141
Db 1144 -----OHSINNVKDK 1153

RESULT 12
T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T18296
R:Guillen, N
submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C:Genetics:
A:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MOW>

Query Match 5.1%; Score 302; DB 2; Length 2139;
Best Local Similarity 20.1%; Pred. No. 0.0013;
```

```
Matches 249; Conservative 206; Mismatches 429; Indels 354; Gaps 54;
QY 59 EGISQRLREYSNKAINKPTKKQYF--SDFINK--SNDLINKDNLID--VESSTKFSQ--- 110
Db 706 EGIRTRMCPYPMRIYSEFLKRYFLGDKINRRASDEKATASTLMRQIIEKYYKKDQCPRA 765
QY 111 KFGDQRYRFTSWVSHQNDPDKINTR-SIRNFMENIQPPIILDDKEKAEFLKSAKOSFAG 169
Db 766 KEGDYRF-----CKTKIFFRVGILAEIEEVREAKI--SEMLGGLQACARAFAA 811
QY 170 IIGNQIRPTDQKPMGVFDESLEK-----EROEAKNEGPEPTGGDWL 208
Db 812 RKHFSVLRBQQAVAVIQRTLSKSWTFRNWMWMLYVKARPMKRRNFEEGKKDKKETE 871
QY 209 DIFLSFIFDKOSSDVKEAINOEPEVPHVQPDIAATTTDIOGLPPEARDLDDERGNF-SKF 267
Db 872 DL-----KKKLAEBELKKREAAENA-----LASATAKTGELEAKIQDLLEKISELSKL 919
QY 268 TLGDMEM-----LDVEGVADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYGGNG 323
Db 920 SAAELDKQELNKLKIEENLEEDKEELKETIDNLKGLDKSLKGEDLVEITELN----- 972
QY 324 GPCARHDMNATVGYKDDQOCNNVATTINVMKNGSGLVIAGGEGKGINNPSEYLYKEDOLT- 382
Db 973 --SQINTLNATYDKKDK-----TIAEQESIDE-----KEDEITK 1005
QY 383 --GSORALSOEETQNKIDPFMEFLAONNAKLDNLSEKEKEKFEITEIKDFQKSKAYLDALG 440
Db 1006 LKGDIKLEEE-----KDDLEQDRADVSA TKDDIA-KLKNITIECED-----A 1048
QY 441 NDRIAFAVSKKDKTHSALITEFGNGDLSYTLKDYGGKAKADKALDREKNVTLQGLSLKHGDMVF 500
Db 1049 KDEIA-----KLEGELEDEEN----- 1064
QY 501 VDYSEKTYTNASKNPNKGVGTNGVSHLEVGNKVAIFNLPDLNLAITSFVRNLE-DK 559
Db 1065 ---KNKDLTNELQOTQLKGETEKSAAQVAATKASDERDTLS-----QNLNENK 1112
QY 560 LTTKGLSQEANKLKDFLSSNKLGVKTLNFNKAVADAKNTGNYDEVKKAQDLEKSLR 619
Db 1113 LTTKNTLTAKD-----LEKKISGLKQDY-EDLEDDKN-----KIEGDLRNAOR 1155
QY 620 KREHLEKEVEKKLE-SKSGNKNMEAKQAQNSOKDEIFALINKANRDRARAIYAQNLAG 678
Db 1156 KIKELDDEITKADYSQYLOKQKEEYESQIAKQBEKEAIGNDVKNKEKTIKEKELEIQS 1215
QY 679 IKRELSO---KLENVNKLKDFDK-----SFDEPKNGKNKDFSKAEETLKALKGS 725
Db 1216 LOEKLDETEVEKEDAEKKKKEIEKEMKALQEBEKNVESKSTKDKKKLEDNLAKDTQKK 1275
QY 726 VKDLGINPEWI-SKVENLNAALNEFKNGKNK-----DFSKV 760
Db 1276 LDDMTADNEKLRKAKARDLEAQLNEVODNIEKAVADAELLNKKKAQSDKELNSLKAELEAL 1335
QY 761 TQAKSLNSVDVITINOK--VTDRVDNLNOAV-----SVAKATGDFSRV----- 803
Db 1336 TRAKSVVESKNKD-SENEKAALSEIDQANEKLNQIADLRKATADLOANEKKAQVEAQ 1394
QY 804 -EQALADLNKFSK--EQL-----AQQAQKNESLNARKKSEI-- 836
Db 1395 RDKLVADNKKMTKTLLEEIKARDEENTYKVENYKVLKRKEADLEENANLDEKDKRMNK 1454
QY 837 -----YQSVKNGVNGTL-----VGNGLSQAQEAATLTK-----NFSDIKKLENAKL 876
Db 1455 EKQVKKLGELEKETKDKLNAAIAEKDSIPTAKKQSDADLEELNKVTVEEHDEVVAKLNTQI 1514
QY 877 GNFNNNNNGLK--NEPIYAKVNNKKKAGQAASLEEPIYAQVAKV-----NA 921
Db 1515 TKLTRDNQSAEELNE-LRSKADKDK-KISELEQVNELESRRPVGTGNADENEIKIRDA 1572
QY 922 KIDRLNQIASGLGVG---QAAGFPLKRHRDKVDLDSKVGLSRNQ-----ELAOKTD 969
Db 1573 QIADLNKALEMKGVQNNQLOATNKEKAKDN-DLTSKIEITENEMKKLENKAKRLEQDQKD 1631
```

```

Qy 970 NLQAVSEAKAGFGNLEOTIDKLDSTKHPNMLWVSAAKKVPASLSAKLDNTATNSHI 1029
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1632 EADKAVSEOTIKRG-LEEEVKL--TTEIOALKFOINA---PSSVAQEEKORLESDI 1684
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1030 RINSNTKGAINKATGMLTQK-----NPEWLKLVNDKIVAHNVGSPVLS 1075
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1685 ---AELKEQLEQERTTAANAERKKIAELDEVFNLEDTNQREKLVAKN--SENDAE 1739
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1076 YDKIGNQKNMKD-----YSDSFKFKSTKLNNAVKDTNS 1108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1740 IDSLEKEKALEIEIKITDDNNKISEIDSLDRKYNA 1777
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
T38077
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38077
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21767
A:Accession: T38077
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1957 <CON>
A:Cross-references: EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SPAC1
A:Experimental source: strain 972h; cosmid c1f3
C:Genetics:
A:Gene: SPDB:SPAC1f3.06c
A:Map position: 1

```

Qy	452	TKUSALITERP	---	NGDLSVTLKDY	GKKADKAL	-	DREKXNVT	LGQSL	KHGDGV	WFDY	SFNKY	508	
			: :			:		:	:	:	:		
Db	846	VNOESKYSEL	KEVNG	SLSLDL	KNLRLSS	NVAIS	DNDQIL	TOLAELSK	-----	NYDSLEQ	899		
			:	:	:	:	:	:	:	:	:		
Qy	509	TNASKXPN	-	KGCV	TNGVSH	---	LEVG	FNKVA	-----	IFN	539		
			:	:	:	:	:	:	:	:	:		
Db	900	ESAQLNSGL	KSLBAEK	QOLUHT	NEELUH	RLDKLT	GKGLK	IESKSD	LCKKL	TARQE	ISN	959	
			:	:	:	:	:	:	:	:	:		
Qy	540	LPOLN	---	NLAITS	FVRRLN	DKLTT	KGSLPQ	EANKLIK	---	FLS	579		
			:	:	:	:	:	:	:	:	:		
Db	960	LKEENMSQ	QAITS	-	VKS	LDLTS	-----	KSSK	LEADIE	HLKNK	VESEVERN	1011	
			:	:	:	:	:	:	:	:	:		
Qy	580	SNKELVGT	LNFNKAVD	-----	AKN	-----	TGN	YDEV	-----	KKAQ	KOLEK	616	
				:	:	:	:	:	:	:	:		
Db	1012	SNRMLD	LKNGENT	ASTQ	TEIKRA	ENDDLO	SKLSV	SEYEN	LLLISS	OTNK	SKSL	1071	
			:	:	:	:	:	:	:	:	:		
Qy	617	SLRKREHL	KEVEK	LESK	SGNKNWE	-----	AKA	ANSQ	-	KDEIF	ALLINKEAN	668	
			:	:	:	:	:	:	:	:	:		
Db	1072	KTQOLKY	TIENK	WQKLDEK	QOORV	LEELUT	SKYK	LGEEN	QAIKDE	LLALPK	KSKKHQ	1131	
			:	:	:	:	:	:	:	:	:		
Qy	669	AIAYAQN	LKIKREL	SJLKEN	YNKNL	KDFK	SDFE	FKN	GKNK	DFSK	ABETL	KALKGS	728
			:	:	:	:	:	:	:	:	:		
Db	1132	CANFVDD	LK	---	EKSD	AEQLT	TNEK	NELIV	LESQ	SNN	EAELVE	ERSDLANRL	1187
			:	:	:	:	:	:	:	:	:		
Qy	729	LGINPEW	ISKVEN	-----	LNAAL	NEFK	NGKNK	---	DFSK	VTOAK	SLENSV	---	772
			:	:	:	:	:	:	:	:	:		
Db	1188	LSBDSNV	IVIS	IRSD	LVRV	DELDT	LTKK	DKDS	LSLTQY	SEVC	QDRD	DLDDL	1247
			:	:	:	:	:	:	:	:	:		
Qy	773	-----	DVIL	INKV	YTDK	-----	VDN	LNAQ	VS	VAKAT	797		
				:	:	:	:	:	:	:	:		
Db	1248	AVSLREL	CTKSE	LDVPS	EILDDN	FVFN	AGNF	SELS	RLTV	LSLE	NLD	FAFNQV	1300
			:	:	:	:	:	:	:	:	:		
Qy	798	GDFSRVB	-----	QALAD	LKNF	SQBQL	AAQAK	NESL	MARK	KS	EIYOSV	RNGV	844
			:	:	:	:	:	:	:	:	:		
Db	1301	NFKK	MELDN	RLTTT	DAEFT	KV	VVAD	EKLQ	HEH	DDWL	IQRGD	EAL	1354
			:	:	:	:	:	:	:	:	:		
Qy	845	NGTLV	GNLSQ	ABAT	---	TL	SKNF	SDTK	KE	LNAK	LNFNNN	---	890
			:	:	:	:	:	:	:	:	:		
Db	1355	R	-----	KEAEM	TENI	HS	LEG	EETK	KEI	TAELS	SRLED	NQ	1406
			:	:	:	:	:	:	:	:	:		
Qy	891	PIYAKV	NKKK	GAGAA	SLSE	PI	YAO	VAKK	VNAK	IDL	RNL	GIAS	950
			:	:	:	:	:	:	:	:	:		
Db	1407	RLKED	VULK	ESLIT	LES	LSNQ	-	RQ	ESS	L	L	DAK	1448
			:	:	:	:	:	:	:	:	:		
Qy	951	DOLSKV	LSRNB	QLAQ	IDNL	NAQVSE	-----	AK	AGF	FG	N	---	997
				:	:	:	:	:	:	:	:		
Db	1449	DDTSR	---	KNSL	MEK	IES	INSS	LDOK	SFEL	AS	AV	EKL	1504
			:	:	:	:	:	:	:	:	:		
Qy	998	KHNP	MLN	WESA	-	KVP	AS	L	SAK	LDNYA	---	TNSH	1046
			:	:	:	:	:	:	:	:	:		
Db	1505	QEAK	ERTQ	VD	EST	IQ	ELD	HEIT	A	S	KNNY	EK	1564
			:	:	:	:	:						

RESULT 14
Tl8477
hypothetical protein C0485w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: Tl8477
R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z18937

A:Accession: T18477

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2523 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1322571; PIDN:CAAL15620.1

C:Genetics:

A:Map position: 3

A:Introns: 148/3

A:Note: C0485W

Query Match 5.0%; Score 294.5; DB 2; Length 2523;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 258; Conservative 177; Mismatches 443; Indels 341; Gaps 61;

QY 30 LKVDNAVASYDP----DQ-KPIVDKNDNRQAFEGISOLREYSKAIKNTKKNQVFS 84
DB 982 LKGENEI-GMDPLMKIDQTNKIVSKVDSGNFNKVDGINFNKTDGSGN-----FNKIDG 1032

QY 85 DFINKSNDL-INKNDLIDVESSTK-SFOKFGDQRYRIFTSWVSHQND--PSKINTSRIN 140
DB 1033 INFNTDGINFNKTDGINFNKTDGINFNKVDNINFNKIDVEKIVDPLPUSHIRTDWIR- 1091

QY 141 FWNIIQPPILDDEKAEFLSAKOSFAGIIGNQIRTDQKFMGVFDESLEKQERQAEKNG 200
DB 1092 -----KKKSELLLSKOGS---IISN-----LDTSHFE----- 1116

QY 201 EPTGGDWLDIFLSFIFDK--KOSSDVKEAINOEPPHVPQDIATTTTIDQIPLPPPEARDLL 258
DB 1117 -----INLSRSEIONEMCKNSVCKOLENKLELEKEIKDEKELNQ-----NEL- 1162

QY 259 DERGNFSKPTLGDMEMLDVEGVADIDPNVKFNOLLIHNNALSSVLMSGHNGIEPEKVSIL 318
DB 1163 -ERSNWS-----IDIE---DLDKD-----LIINKESDIKKYIWDINKDNYMMI 1203

QY 319 YGNGGPGARHDWNAVTVGYKQDQGNVATII---NVHMKNGSLVIAGGEGINNPSFYL 375
DB 1204 YQDN-----KCGRRKMKWISQNKLLIKKRIKRNHE-----KKRKIRFFFKL 1245

QY 376 YKEDQITGSORALSQEQEIQNKIDFMEFLAQNNAKLDNLSEKEKEFRTEIKDFQKDSK-- 433
DB 1246 YK-----RNDTHKKLRPFRVHRVDVKLDNLNDK--TVMLKNEIRVDKGEDRGE 1292

QY 434 -AYLDALGNDRIA--FVSKKOTKHSALITEFGNGDLSYTLKDYGKKADKALDREKNVTIQ 490
DB 1293 DYVYDFLNKDNMGNMKNKKNVKNVKNVNN-----VKDV-----KNVNVKNVNV 1341

QY 491 GSKLDGVFVVDYSNFKYT---NASKNPKNKGVTVNGVSHLEVGNKVAIFNLPLD----- 543
DB 1342 NNVN-----VKDVKNMEHIDKYNKKEVMIKKGESNNVPKHEKHNKNKNCNY-DLGMHS 1396

QY 544 --NLAITS-----FVRNLEDKLTGSLSPQEAENKLIKDFLSSNKELVGKTLNFKAVA 596
DB 1397 LQNRITITSEVSSKFLCKRMKNYF--DKSNNSIEIHKISASNIFRITMVCASNK-----G 1450

QY 597 DAKNTGNYDEV-----KKAQKDLEKSLRKRHELEKEVEKLEKSKGNKNK 641
DB 1451 ENKNGNINNYKGPATKALVKNKLFISKKESKRAITSSKKRDDNINVIKKTINTPS---QK 1507

QY 642 MEAKAQAQSQDEIPALLNKEANRDARAIAQNLIK--GIRKRESDKLENNVKNLKD--- 696
DB 1508 VSEKRNNNNNNNVLDGNKKNKNDE---LFTKEIKKSTISQKQKKGNEGNTKTHKDNIN 1564

QY 697 -FKSFEDEF-----KNGKNKDFSKAETLALKGSKVDLGINPEWISKVENLNA 744
DB 1565 ILNEDVDHFKOPSLRLEVTKKNNKN-----KNNKNKKFNNDYNNNNNNNNNS 1612

QY 745 ALNEFKNGKNKDFSK---VTOAKSDL-----ENSVKD--VIINQKVTDKVDNLNQAQVSAK 795
DB 1613 --NDEEYKEEHIAITNEIVRESESDLYVSCDEGCVKNGDIYNMEIINNVDNIHK---IDK 1667

QY 796 ATGDFSRVQALADLKNFSEQL-----AQAQAKNESLNARKKSEIYQSVKNG 843
DB 1668 KGNIDIDYKDKS---LENNKKQKGLIKLTPSSYTEKEKEKEKERRKKNNIYIP--- 1720

QY 844 VNGTLVGNCLSQAEATLTSKNESDIKKELNAKLGNFNNNNNGLNNEPIYAKV---NKK 899
DB 1721 -----LTIASRTVAIPIINDTTKDVKNKLHLTKRNTC-----VTYCNVDNIQNK 1765

QY 900 KAGQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVGOAGFPKLRHDKVDDLSKVGLS 959
DB 1766 KKGD-----DKKNIKRQHVGLKFLDEM--SAMFEKKKKIKKDDINKKEDI 1810

QY 960 RNQELAQIDNLNQAVSEAKAGFFGNLEQTIDKLDKSTKHNPMLNWLWESAKKVPASLSAK 1019
DB 1811 NKDDINKKDNINK-----KNDINKKDDINKKDDINKKDDINKKDDIN---K 1854

QY 1020 LDNYATNSHIRINSNI-----KNGAINEKATGMLT-----QKNPEWLKLVNDKIYVAH 1066
DB 1855 KNNYNNNSN---NNNVVKFKSTHQNEEKIKGNITVIRNKLKDKGKENIGLKKKIERK 1911

QY 1067 NVGSVPLSEYDK-IGFNOKNMKDYSDSFKFSKTLNNAVKTNSGFTQFLTNFASASYC 1125
DB 1912 NTTTTATKHHNIDIIKKKNEKE-----NKMKN-----SKFQC 1945

QY 1126 LARENAE-HGINKVNTKGG 1143
DB 1946 LKNKGTQIENKKNMIISG 1964

RESULT 15

A64465

hypothetical protein MJ1322 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: A64465

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Black

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannas

A:Reference number: A64300; MUID:96337999

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1005 <BUL>

A:Cross-references: GB:067572; GB:L77117; NID:g1591958; PIDN:AAB99331.1; PID:g1591962

C:Genetics:

A:Map position: REV1273394-1270377

C:Superfamily: hypothetical protein MJ1322

Query Match

Best Local Similarity 5.0%; Score 293.5; DB 2; Length 1005;

Matches 183; Conservative 134; Mismatches 278; Indels 251; Gaps 41;

QY 338 KDOQGNVATTIIVHMKNGSLVIAGGEGINNPSFYLYKEDQLTGSQRALSQEQEIQNKI 397
DB 6 KEIRNNPKSHVNSRIKEKGIIVGSGSKSIF-----EAV 45

QY 398 DFMEFLAQNNAKLDNLSEKEKEFRTEIKDFQKOSKAYLDALGNDRIFAQVSKKDTKHSAL 457
DB 46 PFALFGAGSNFYDTIITKGGKVVVEL-DPEVNGNNY-----KI 84

QY 458 ITEFGNDLSYTLKDYGKKADKALDREKNVTLOGSLKHGDMVFDYSNFK-YTNASKNPN 516
DB 85 IREYDSG-----RGAKL-----YKNGKPYATTISAVN 112

QY 517 KGVGVTVNGVSHLEVGNKVAIFNLPLDNLNLAITSFVRNLEDKLTTKGLSPQEAENKLIK 576
DB 113 KAVNEILGVD-----RNMFLNSIYIKQGE---IAKFLSLKPSKEL--- 149

QY 577 FLSSNKELVGKTLNPNKAVADAKNTGNYDEVKKAQKLEKSLRKRHELEKEVEKLEKSKS 636

Db 150 -----ETVAKLLGID-----: || :| : || :|| :||
Qy 637 GNKNMEAKAQAQNSQKDEIFALINKEANDARATAYAQNLKGIKRELSKLENVKNLKD 696
Db 184 GELNYKE-----NYEKE-----LKNKMSQLEEK-----NKK--LWEINDKLNKIKKEFED 226
Qy 697 FDKSPDEFKNGK---NKDFSKAEETLKALGKSVKDLGINPEWISKV---ENLNAALNEF 749
Db 227 IEKLEFNEWENKLLYEKFINKLEERKRALELNQELKILEYDLNTVVEARETLNRHKDEY 286
Qy 750 KNGKN-----KDFSKVTQ-----AKSDLENSVKDVIINQKVTDKVDN 786
Db 287 EKYKSLVDEIRKIESRLRELKSHVEDYKLTQKLEIKGDIE-KLKEFINKSKYRDDIDN 345
Qy 787 LQAVSVAKATGDFSRVE---QALADLNFSKE-----QLAQQAQK--NESLNARKK 833
Db 346 LDTLLN--KIKDEITERVETIKDLLLEELKNLNEEIEKIEKYKRICECKEYVEKYLEEEK 403
Qy 834 SEIYQSVNGVNGTLVGNGLSQAEATTLSKNPFSDIKKELNAKLGNNNNNNGLKNEPIY 893
Db 404 AVEYNKL-----TLEYITLQ-EKKSIEKNINDLETRINKLLEETKNIDIESIENS--L 454
Qy 894 AKVNNKKAGQAASLEPIYAQVAKK---VNAKIDRLNOIASGL-GVVGQAAGFPLKRHDK 949
Db 455 KEIEEKKVLENLQKEKI--ELNKKLGEINSEIKRLKILDELKEVEGKC---PLCK-TP 508
Qy 950 VDDLKSVGLSRNQELAQ-----KIDNLQAVSEAKAGFFGNLEQITDKL-----KDSTK 998
Db 509 IDENKKMELI-NOHKTQLNKNYTELEENKKIRE-----IEKDIKLEKKEIDKEENL 559
Qy 999 HNPMLWVESAKKVPASLSAKLDNYATNSHIRINSINNGAINERKATGMLTQ-----1050
Db 560 KTLKTYLEKOSQI-ELELKLKNYKEQLD-EINKKISNYVINGKVPDVEILEDIKSQLNK 617
Qy 1051 -KN-----PEWLKLYNDKIVAHNGSVPLSEYDKI--GFNOKNMKDYSDSPKFSTKL 1099
Db 618 FKNFYNQYLSAVSYLNSVDEEGIRNRI-----KEIENIVSGWNKEKREELNKLREDERE 672
Qy 1100 NNAVKD 1105
Db 673 INRLKD 678

Search completed: February 13, 2002, 14:03:48
Job time: 3047 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 14:02:21 ; Search time 68.44 Seconds

(without alignments)

614.473 Million cell updates/sec

Title: US-09-360-685A-5

Perfect score: 5867

Sequence: 1 MTNETIDQPPQTEAENFPOQ.....RENAERGIKNVNTKGGFQKS 1147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5867	100.0	1147	1 CGAL_HELPY	P80200 helicobacte
2	5395.5	92.0	1186	1 CAGA_HELPY	P55980 helicobacte
3	5270.5	89.8	1182	1 CGA2_HELPY	P55746 helicobacte
4	5086	86.7	1167	1 CAGA_HELPY	P921t1 helicobacte
5	314	5.4	1790	1 US01_YEAST	P25386 saccharomyc
6	307	5.2	1251	1 RBP2_PLAVB	Q00799 plasmodium
7	301	5.1	1957	1 YD86_SCHPO	Q10411 schizosacch
8	291	5.0	2230	1 G0G4_HUMAN	Q13439 homo.sapien
9	289	4.9	2663	1 CENE_HUMAN	Q02224 homo.sapien
10	284	4.8	1875	1 MLP1_YEAST	Q02455 saccharomyc
11	279	4.8	1679	1 Y109_YEAST	P40457 saccharomyc
12	276.5	4.7	2869	1 RBP1_PLAVB	Q00798 plasmodium
13	276	4.7	1928	1 MYS1_YEAST	P08964 saccharomyc
14	272.5	4.6	2022	1 ANP1_ONCVO	P21249 onchocerca
15	268	4.6	1208	1 PCP1_SCHPO	Q92351 schizosacch
16	267	4.6	1727	1 ALM1_SCHPO	P47460 mycoplasma
17	267	4.6	1805	1 HMW2_MYCEE	P92utk5 schizosacch
18	266	4.5	1938	1 MYS2_CAEEL	P02567 caenorhabdi
19	265	4.5	1966	1 MYSB_CAEEL	P02566 caenorhabdi
20	262	4.5	2116	1 MYS2_DICDI	P08799 dictyosteli
21	260.5	4.4	2704	1 BPA1_HUMAN	Q03001 homo.sapien
22	255	4.3	1935	1 MYH7_RAT	P02564 rattus.norv
23	254.5	4.3	1312	1 RA50_YEAST	P12753 saccharomyc
24	252	4.3	1938	1 MYH4_RABIT	P28641 oryctolagus
25	251.5	4.3	1959	1 MYH9_CHICK	P14105 gallus.gall
26	250	4.3	1972	1 MYHB_HUMAN	P35749 homo.sapien
27	249.5	4.3	1938	1 MYS_AEQIR	P24733 aequipecten
28	249	4.2	1630	1 MSP1_PLAFW	P04932 plasmodium
29	249	4.2	1639	1 MSP1_PLAFW	P04933 plasmodium
30	248.5	4.2	1972	1 MYHB_RABIT	P35748 oryctolagus
31	247.5	4.2	1972	1 MYHB_MOUSE	O08638 mus.musculu
32	247	4.2	1934	1 MYH7_MESAU	P13540 mesocricetu
33	246.5	4.2	1939	1 MYH1_HUMAN	P12882 homo.sapien

34	245.5	4.2	1939	1 MYH6_MESAU	P13539 mesocricetu
35	245	4.2	1935	1 MYH7_PIG	P79293 sus.scrofa
36	245	4.2	1938	1 MYH6_MOUSE	Q02566 mus.musculu
37	245	4.2	1939	1 MYH6_HUMAN	P13533 homo.sapien
38	243.5	4.2	2349	1 TPR_HUMAN	P12270 homo.sapien
39	243	4.1	1818	1 HMW2_MYCPN	P75471 mycoplasma
40	243	4.1	2104	1 MYS3_SCHPO	O14157 schizosacch
41	242	4.1	1427	1 REST_HUMAN	P30622 homo.sapien
42	240.5	4.1	1938	1 MYH6_RAT	P02563 rattus.norv
43	240	4.1	1935	1 MYH7_HUMAN	P12883 homo.sapien
44	238.5	4.1	1960	1 MYH9_HUMAN	P35579 homo.sapien
45	238.5	4.1	2710	1 TOXA_CLODI	P16154 clostridium

ALIGNMENTS

RESULT 1

ID	CGAL_HELPY	STANDARD;	PRT;	1147 AA.
AC	P80200;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN).			
GN	CAGA OR CAL.			
OS	Helicobacter pylori (Campylobacter pylori).			
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
OC	Helicobacter.			
OX	NCBI_TaxID=210;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CCUG 17874 / NCTC 11638;			
RX	MEDLINE=93296225; PubMed=8516329;			
RA	Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,			
RA	Macchia G., Massone A., Papini E., Xiang Z., Figura N.,			
RA	Rappuoli R.;			
RT	"Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=CCUG 17874 / NCTC 11638;			
RX	MEDLINE=97121442; PubMed=8962108;			
RA	Censini S., Lange C., Xiang Z., Crabtree J., Ghiara P.,			
RA	Borodovsky M., Rappuoli R., Covacci A.;			
RT	"cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).			
[3]				
SEQUENCE OF	181-90;314-28;366-77;420-30;494-506;661-77;900-14;1062-77.			
Herrmann V., Herrmann J., Kist M.;				
Submitted (APR-1993) to the SWISS-PROT data bank				
!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT, OR FUNCTION OF THE CYTOTOXIN.				

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				

EMBL; X70039; CAA49633.1; -				
EMBL; AF282853; AAC44706.1; -				
HSSP; P02937; IMLP.				
Antigen.				
DOMAIN	880	885		POLY-ASN.
CONFLICT	320	320		G -> A (IN REF. 3).
CONFLICT	325	325		P -> F (IN REF. 3).
CONFLICT	328	328		R -> K (IN REF. 3).

```
FT CONFLICT 426 426 K -> E (IN REF. 3).
FT CONFLICT 429 429 Q -> E (IN REF. 3).
FT CONFLICT 673 675 AON -> TED (IN REF. 3).
FT CONFLICT 901 901 A -> T (IN REF. 3).
FT CONFLICT 903 903 Q -> E (IN REF. 3).
FT CONFLICT 907 907 L -> P (IN REF. 3).
FT CONFLICT 910 910 P -> R (IN REF. 3).
FT CONFLICT 914 914 Q -> E (IN REF. 3).
FT CONFLICT 1072 1072 P -> S (IN REF. 3).
FT CONFLICT 1074 1074 S -> D (IN REF. 3).
SQ SEQUENCE 1147 AA; 128013 MW; AB92770835F68490 CRC64;

Query Match 100.0%; Score 5867; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 5.5e-197;
Matches 1147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTNETIDQOQTEAAFPNPOQFINNQLQVAFKVDNAVASYDPOKPIVDKNDNRNROAFEG 60
DB 1 MTNETIDQOQTEAAFPNPOQFINNQLQVAFKVDNAVASYDPOKPIVDKNDNRNROAFEG 60
QY 61 ISQLEEYSNKAIKNPTKKNQYFSPDFINKSNDLKNKNDLIDVESSTKSPQKFCDOYRIF 120
DB 61 ISQLEEYSNKAIKNPTKKNQYFSPDFINKSNDLKNKNDLIDVESSTKSPQKFCDOYRIF 120
QY 121 TSWVSHQNDPSKINTSRNFMENIIQPPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180
DB 121 TSWVSHQNDPSKINTSRNFMENIIQPPILDDKEAEFLKSAKOSFAGIIGNQIRTDQ 180
QY 181 KFMGVFDESLEKQEAKEKNGPTGGDWLDFISFTFDKQSSDVKEAINQEPVPHVQPD 240
DB 181 KFMGVFDESLEKQEAKEKNGPTGGDWLDFISFTFDKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTDIOGLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
DB 241 ATTTDIOGLPEARDLDERGNFSGFTLGDMEMLDVEGVADIDPNYKFNOLLHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWNTATVGYKDOQGNVATIIINVHMKNGSLV 360
DB 301 SVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWNTATVGYKDOQGNVATIIINVHMKNGSLV 360
QY 361 IAGGEGKINNPFFLYKEDQLTGSORALSQEIQNKIDPMEFLAONAKLDNLSEKEK 420
DB 361 IAGGEGKINNPFFLYKEDQLTGSORALSQEIQNKIDPMEFLAONAKLDNLSEKEK 420
QY 421 FRTKIDFODKSKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGGKADKA 480
DB 421 FRTKIDFODKSKAYLDALGNDRIFAIVSKKDTKHSALITEFGNGDLSYTLKDYGGKADKA 480
QY 481 LDREKNVTIQQSLKHDGVNFVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKNVAIFNL 540
DB 481 LDREKNVTIQQSLKHDGVNFVDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKNVAIFNL 540
QY 541 PDLNNIATSFVRNLEDKLTTKGLSPQBAKLIKDFLSSNKELVCKTLNPNKAVADAKN 600
DB 541 PDLNNIATSFVRNLEDKLTTKGLSPQBAKLIKDFLSSNKELVCKTLNPNKAVADAKN 600
QY 601 TGNVDEVKKAQDLKSLKREHLEKEVEKFLSKSGNKNKWEAKAQAQNSQKDEIFALIN 660
DB 601 TGNVDEVKKAQDLKSLKREHLEKEVEKFLSKSGNKNKWEAKAQAQNSQKDEIFALIN 660
QY 661 KEANDARAIAAQNKGKIKRELSKLENVNKNLKDQDKSFDFKNGKNKDFSKAEETLK 720
DB 661 KEANDARAIAAQNKGKIKRELSKLENVNKNLKDQDKSFDFKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENIENALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINOKV 780
DB 721 ALKGSVKDLGINPEWISKVENIENALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINOKV 780
QY 781 TDKVDNLNQAVSAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
DB 781 TDKVDNLNQAVSAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
```

```
QY 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNNNNNNGLKNEPIYAKVNKK 900
DB 841 KNGVNGTLVGNGLSQAEATTLTKNFSDIKKELNAKLGNNNNNNGLKNEPIYAKVNKK 900
QY 901 AGQAASLEPIYAQVAKVNAKIDRLNQITASGLGVVGOAAGPPLKRHRDKVDLDSKVGLSR 960
DB 901 AGQAASLEPIYAQVAKVNAKIDRLNQITASGLGVVGOAAGPPLKRHRDKVDLDSKVGLSR 960
QY 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHNPMLWVESAQKVPASLSAKL 1020
DB 961 NOELAQKIDNLNOAVSEAKAGFFGNLEQITDKLKDSTKHNPMLWVESAQKVPASLSAKL 1020
QY 1021 DNYATNSHIRINSNIKNGAINEKATGMLTKQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIG 1080
DB 1021 DNYATNSHIRINSNIKNGAINEKATGMLTKQKNPEWLKLVNDKIVAHNVGSVPLSEYDKIG 1080
QY 1081 FNQKNMKDYSDSFESTKLNNAVKTNSGFTQFLNFASTASYCYCLARENAEHGKNVNT 1140
DB 1081 FNQKNMKDYSDSFESTKLNNAVKTNSGFTQFLNFASTASYCYCLARENAEHGKNVNT 1140
QY 1141 KGGFOKS 1147
DB 1141 KGGFOKS 1147

RESULT 2
CAGA_HELPY
ID CAGA_HELPY STANDARD; PRT; 1186 AA.
AC P55980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN)
DE (CAG PATHOGENICITY ISLAND PROTEIN 26).
GN CAGA OR CAI OR CAG26 OR HP0547.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilave A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathley L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RL Nature 388:539-547(1997).
RC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
OR FUNCTION OF THE CYTOTOXIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000569; AAD07614.1;
DR TIGR; HP0547;
KW Antigen; Complete proteome.
FT DOMAIN 247 250 POLY-THR.
FT DOMAIN 883 889 POLY-ASN.
SQ SEQUENCE 1186 AA; 132386 MW; B05C3F2CCC4444F4 CRC64;
```


Query Match 92.0%; Score 5395.5; DB 1; Length 1186;
Best Local Similarity 89.9%; Pred. No. 1.3e-180;
Matches 1068; Conservative 26; Mismatches 51; Indels 43; Gaps 5;

QY 1 MTNETIDQ-----QPOTEAAFPQOFINNLOVAFLKVDNAVASYDDPKPIVDKNDNR 55
DB 1 MTNETIDQTRTDPQVQSOTAFDPQFINNLOVAFLKVDNAVASFDDPKPIVDKNDNR 60

QY 56 QAFEGISQLREYSNKAINKPTKKQYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQ 115
DB 61 QAFDGISQLREYSNKAINKPTKKQYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQ 120

QY 116 RYRIFTSWVSHONDPSKINTRISIRNPMENIIQPPILDDKEAEFLKSAKQSFAGIIGNQ 175
DB 121 RYQIFTSWVSHONDPSKINTRISIRNPMENIIQPPIDDKAEFLKSAKQSFAGIIGNQ 180

QY 176 IRTDQKFMGVFDESILKEROEAKNEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPH 235
DB 181 IRTDQKFMGVFDESILKEROEAKNEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPH 240

QY 236 VQPDIAATTTDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLTH 295
DB 241 VQPDIAATTTDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLTH 300

QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDMNATVGYKDOQGNVATIIIVHMK 355
DB 301 NNALSSVLMGSHNGIEPEKVSLLYAGNGFGDKHDMNATVGYKDOQGNVATIIIVHMK 360

QY 356 GSGLVITAGKEGINNPSFYLKEDDLTGSRALSOEIQNKIDFMEFLAQNKLIDNLSE 415
DB 361 GSGLVITAGKEGINNPSFYLKEDDLTGSRALSOEIRKVDMEFLAQNKLIDNLSE 420

QY 416 KEKEKFRTEIKDFQKDSKAYLDALGNDRIFAFVSKKDKHSAITFEFGNDLSYTLKDYCK 475
DB 421 KEKEKFRTEIKDFQKDSKAYLDALGNDRIFAFVSKKDKHSAITFEFGNDLSYTLKDYCK 480

QY 476 KADKALDREKNVTLOGSLKHGDMFVDSNFYTNASKNPNKGVGTNGVSHLEVGFNFKV 535
DB 481 KADKALDREKNVTLOGSLKHGDMFVDSNFYTNASKNPNKGVGTNGVSHLEVGFNFKV 540

QY 536 AIFNLPDLNLNLAITFVRNRLNEDLTKGLSPQERANKLIDFLSSNKELVGLTKLNFNAV 595
DB 541 AVFNLPDLNLNLAITFVRNRLNENKLTAKGLSQERANKLIDFLSSNKELVGLTKLNFNAV 600

QY 596 ADAKNTGNYDEVKKAQKLEKSLRKEHLEKEVEKLESKGNKNKMEAKAQAANSOKDEI 655
DB 601 AEAKSTGNYDEVKKAQKLEKSLRKEHLEKEVEKLESKGNKNKMEAKAQAANSOKDEI 660

QY 656 FALINKEANRDARAAYAQNLKIGIKRELSDKLENVNNKLDKDFKSPDFKNGKNDKFSKA 715
DB 661 FALINKEANRDARAAYAQNLKIGIKRELSDKLEKISDKLDKFSKSPDFKNGKNDKFSKA 720

QY 716 EETLKALGSVKDLGINPWIWISKVENLNAALNFKNGKNKDFSKYTOAKSDLENSVKDVI 775
DB 721 EETLKALGSVKDLGINPWIWISKVENLNAALNFKNGKNKDFSKYTOAKSDLENSVKDVI 780

QY 776 INQVTDKVDNLNQAVSAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSE 835
DB 781 INQVTDKVDNLNQAVSAKMGDFSRVQVLADLKNFSKEQLAQQAQKNEDFNTCKNSE 840

QY 836 IYQSVKNGVNGTLVGNLSQAATLTSKNFSDIKKELNAKLNNNNNNGLKN--EPIY 893
DB 841 LYQSVKNSYNTLVGNLSGIEATLAKNFSDIKKELNEKFNFN--NNNNGLKNSTEPIY 899

QY 894 AKVNNKKAGQAASLEPIYAQVAKVNAKIDRLNOITASGLGVVGOAAGPPLKRHKVDL 953
DB 900 AKVNNKKTQVASEPIYITQVAKVNAKIDRLNOITASGLGVVGOAAGPPLKRHKVDL 959

QY 954 SKVGL-----SRNQLAOKIDNLNQAVSEAK 979
DB 960 SKVGLSASPEPIYATIDDLGGPPPLKRHKVDLDSKVGSRNQELAQKIDNLNQAVSEAK 1019

QY 980 AGFFGNLEQITDKLKDSTKHNPMLWVESAKKVPASLSAKLDNYATNSHIRNSNIKCA 1039
DB 1020 AGFFGNLEQITDKLKDSTKKNYNNLVESAKKVPASLSAKLDNYATNSHIRNSNIQCA 1079

QY 1040 INEKATGMTOKNPEWLKLVNDKIVAHNNVGSVPLSEYDKIGFNQKNMKYSDSFKESTKL 1099
DB 1080 INEKATGMTOKNPEWLKLVNDKIVAHNNVGSVLSYDKIGFNQKNMKYSDSFKESTKL 1139

QY 1100 NAVKDNGSTGFTQFLTNAFSTASYCYCLARENAEHGKKNVNTKGGFOKS 1147
DB 1140 NNAVXDKSGFTHFLANAFT-GYYCLARENAEHGKKNVNTKGGFOKS 1186

RESULT 3
CGA2_HELPY
ID CGA2_HELPY STANDARD; PRT; 1182 AA.
AC P5746;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN).
GN CAGA OR CAI.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 53726 / 84-183;
RX MEDLINE=93239281; PubMed=8478069;
RA Tumuru M.K.R., Cover T.L., Blaser M.J.;
RT "Cloning and expression of a high-molecular-mass major antigen of Helicobacter pylori: evidence of linkage to cytotoxin production.";
RL Infect. Immun. 61:1799-1809(1993).
CC !- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L11714; ; NOT_ANNOTATED_CDS.
DR HSP: P02937; 1MLP.
KW Antigen.
FT DOMAIN 878 885 POLY-ASN
SQ SEQUENCE 1182 AA; 131503 MW; C916817E2EE57BB4 CRC64;

Query Match 89.8%; Score 5270.5; DB 1; Length 1182;
Best Local Similarity 88.3%; Pred. No. 2.8e-176;
Matches 1046; Conservative 33; Mismatches 66; Indels 39; Gaps 6;

QY 1 MTNETIDQOPTEAAFPNQOFINNLOVAFLKVDNAVASYDDPKPIVDKNDNRQAFEG 60
DB 1 MTNETIDQOPTEAAFPNQOFINNLOVAFLKVDNAVASYDDPKPIVDKNDNRQAFEG 60

QY 61 ISQLEEYSNKAINKPTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQYRIF 120
DB 61 ISQLEEYSNKAINKPTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDQYRIF 120

QY 121 TSWVSHONDPSKINTRISIRNPMENIIQPPILDDKEAEFLKSAKQSFAGIIGNQIRTDQ 180
DB 121 TSWVSHONDPSKINTRISIRNPMENIIQPPIDDKAEFLKSAKQSFAGIIGNQIRTDQ 180

QY 181 KFMGVFDESILKEROEAKNEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPHVQPD 240
DB 181 KFMGVFDESILKEROEAKNEPTGGDWLDFLSPFDKQSSDVKEAINOEPVPHVQPD 240

QY 241 ATTTDDIOGLPPEARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLHNNALS 300

```

Db 241 ATSTTHIQGLPPSRDLDLDERGNSKFTLGDMEMLDVEGVADMDPNKFNQLLHNNTLS 300
QY 301 SVLMGSHNGTEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATIIIVHMKNGSGIV 360
Db 301 SVLMGSHDGEPEKVSLLYAGNGGFGAKHDWATVGYKDOQGNVATIIIVHMKNGSGIV 360
QY 361 IAGGEGKINNPSPYLYKEDQLTGSGORALSQEEIQNKIDFMFLAONNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPSPYLYKEDQLTGSGORALSQEEIQNKIDFMFLAONNAKLDNLSEKEK 420
QY 421 FRTEIKDFQKSKAYLDALGNDRIFAIVSKKDYKHSALITEFNGDLSYTLKDYGKKADKA 480
Db 421 FRNEIKDFQKSKPYLDALGNDRIFAIVSKKDPKHSALITEFNGKGLSYTLKVMGKKQIK 480
QY 481 LDREKNTVLQGLSKHDGVNFVDYSNFKYTNASKPNKNGVGTNGVSHLVEGKNAIFNL 540
Db 481 LDREKNTVLQGLSKHDGVNFVSNFKYTNASKPNKNGVGTNGVSHLVEGKNAIFNL 540
QY 541 PDNLNLAITSVRRNLEDKLTGKLSPOBANKLIKDFLSSNKELVKTILNFKAVADAKN 600
Db 541 PNLNLAITSVRRDLEDKLIAGLSPQEBANKLVKDFLSSNKELVCKALNFENKAVAEAKN 600
QY 601 TGNVDEVKKAQDLKSLKREHLEK-EYEKLEKSGKNKMEAKAQAANSOKDEIFALI 659
Db 601 TGNVDEVKKAQDLKSLKREHLEKGDVAKNLEKSGKNKMEAKAQAANSOKDEIFALI 660
QY 660 NKEANRARAIAVAQNLKGIKRELSDKLENVKNLKDDEKSPDEPKNGKNDKFSKAEETL 719
Db 661 NKEANRARAIAVAQNLKGIKRELSDKLENINRDLKDFSKDFGKNGKNDKFSKAEETL 720
QY 720 KALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTQAKSDLENSKVDLIINQK 779
Db 721 KALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTQAKSDQENSIDVLIINQK 780
QY 780 VTDKVDNLNQAVSVAKATGDFSRVEQALADLNKFSKEQLAQQAQKNESLNARKKSEIYOS 839
Db 781 ITDKVDNLNQAVSVAKIACDFSGVEQALADLNKFSKEQLAQQAQKNESFNV-GKSEIYOS 839
QY 840 VKNVNGTLVGNGLSQAEATTLTKSNFSDIKKELNAKLGNFNNNNGLKN--EPIYAKVN 897
Db 840 VKNVNGTLVGNGLSGIEATLAKNFSDIKKELNEKFKNFNNNNGLKNGGEPYIAQVN 899
QY 898 KKKAGQAASLEPIYIAQVAKVNAKIDRLNQIA-SGLGVVQGA----- 939
Db 900 KKKTGQVASEPIYIAQVAKVTKIDQLNQATSGFGVGQAGPPLKRDHVEDLSKVG 959
QY 940 -----AGPPLKRDHVKVDLSKVGLSRNOELAQKIDNLNQAVSEAKAGFF 983
Db 960 RSVSPPEPIYATIDDLGGSPPLKRDHVKVDLSKVGLSRNOELTQKIDNLNQAVSEAKAGFF 1019
QY 984 GNLQETIDKLDKSTKHNPNLWAVESAKKVPASISAKLDNATNSHTRINSNIKNGAIN 1043
Db 1020 GNLQETIDKLDKFTKNNPNLWAVESAKKVPASISAKLDNATNSHTRINSNIKNGAIN 1079
QY 1044 ATCMLTQKNPEWLKLVNDKIVAHNVGSPVLSYDKIFGKQNMKDYSDSFKESTKLNNAV 1103
Db 1080 ATGTERQKNPEWLKLVNDKIVAHNVGSPVLSYDNIYSGKQNMKDYSDSFKESTKLNNAV 1139
QY 1104 KDTNSGTOFLNAPSTASYCLARENAEHBGKTNVNTKGGFGKS 1147
Db 1140 KDTKSFTOFLANAFST-GYISWARENABHGKTNVNTKGGFGKS 1182

```

```

RESULT 4
CAGA_HELPJ
ID CAGA_HELPJ STANDARD: PRT: 1167 AA.
AC Q9ZLT1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN (120 KDA PROTEIN)
DE (CAG PATHOGENICITY ISLAND PROTEIN 26).

```

```

GN CAGA OR CAI OR CAG26 OR JHP0495.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC - FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001483; AAD06073.1;
KW Antigen; Complete proteome.
FT DOMAIN 246 249 POLY-THR.
FT DOMAIN 882 889 POLY-ASN.
SQ SEQUENCE 1167 AA; 129729 MW; FD5E86B81CEBD0F2 CRC64;

```

```

Query Match 86.7%; Score 5086; DB 1; Length 1167;
Best Local Similarity 84.1%; Pred. No. 6.8e-170;
Matches 996; Conservative 60; Mismatches 73; Indels 56; Gaps 4;

QY 1 MTNETIDQOQTEAFANFPOQFINNQVAFKVDNNAVSYDDQKPIVDKNDNRQAFEG 60
Db 1 MTNEALNQOQTEAFANFPOQFINNQVAFKVDNNAVSYDDQKPIVDKNDNRQAFEG 60
QY 61 ISQLEESYNSKAIKNPTKKNQYFSDFIKNSDLINKNDLIDVESSTKFSQKEDQYRIF 120
Db 61 ISQLEEFANKAIKNTKKNQYFSSFIKNSDLINKNDLIDGTSSIKFSQKEDQYRIF 120
QY 121 TSWSHQNDPDKINRINFMENIIPDLDKEAEPLKSAKQSAFAGIIGNQIRTOQ 180
Db 121 MNWVSHQNDPDKINRINFMENIIPDLDKEAEPLKSAKQSAFAGIIGNQIRTOQ 180
QY 181 KPMGVFDESLEKROEAKNGE---PTGGDWLDIFLSFDFDKKQSSDVKEAINOEPVPHV 236
Db 181 KPMGVFDESLEKROEAKNGE---PTGGDWLDIFLSFDFDKKQSSDVKEAINOEPVPHV 240
QY 237 QPDIAATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHN 296
Db 241 QPDVATTTDIOGLPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHN 300
QY 297 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATIIIVHMKNG 356
Db 301 NALSSVLMGSHNGIEPEKVSLLYGGNGGPGCARHDWATVGYKDOQGNVATIIIVHMKNG 360
QY 357 SGLVIAGGEGKINNPSPYLYKEDQLTGSGORALSQEEIQNKIDFMFLAONNAKLDNLSEK 416
Db 361 SGLVIAGGEGKINNPSPYLYKEDQLTGSGORALSQEEIQNKIDFMFLAONNAKLDNLSEK 420
QY 417 EKEKFTETKDFQKSKAYLDALGNDRIFAIVSKKDYKHSALITEFNGDLSYTLKDYGKK 476
Db 421 EKEKFTETKDFQKSKAYLDALGNDRIFAIVSKKDYKHSALITEFNGDLSYTLKDYGKK 480
QY 477 ADKALDREKNTVLQGLSKHDGVNFVDYSNFKYTNASKPNKNGVGTNGVSHLVEGKNA 536
Db 481 ADKALDREKNTVLQGLSKHDGVNFVDYSNFKYTNASKPNKNGVGTNGVSHLVEGKNA 540

```

```
Qy 537 IFNLPDLNLAITSFVRNLEDKITTKGLSPQEQANKLIKDFLSSNKELVKGTLNFKAVA 596
Db 541 VFNLPNLAITSFVRQDLEDKIAGKLSQEQANKLVKDFLSSNKELVKGALNFKAVA 600
Qy 597 DAKNTGNYDEVKKAQDKLEKSLRKEHLEKVEKLESKGNKNKMEAKAQSOKDEIF 656
Db 601 EAKNTGNYDEVKKAQDKLEKSLRKEHLEKVEKLESKGNKNKMEAKAQSOKDEIF 660
Qy 657 ALINKEANRDAIAYAOGLKRELSDKLENYKNKIKDFDKFDFEFGKNGKDFSKAE 716
Db 661 ALINKEANRDAIAYAOGLKRELSDKLENYKNKIKDFDKFDFEFGKNGKDFSKAE 720
Qy 717 ETLKALGSKVDGLINPEWISKVENLAALNFKNGKNGKDFSKVTOAKSDLENSVKDVII 776
Db 721 ETLKALGSKVDGLINPEWISKVENLAALNFKNGKNGKDFSKVTOAKSDLENSIKDVII 780
Qy 777 NOKYTKVDNLNOAVSVAKATGDFSRVEQALADLNKFSKEQALAOQAKNESLNARKKSEI 836
Db 781 NOKYTKVDNLNOAVSVAKATGDFSRVEQALADLNKFSKEQALAOQAKNESLNARKKSEI 840
Qy 837 YOSYKNGVNGTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNFNNNNGLNKNEPIYAKV 896
Db 841 YOSYKNGVNGTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNFNNNNGLNKNEPIYAKV 894
Qy 897 NKKKAGQAASLEPIYAOVAKKYNKADIRLNOIASGLGVGVQAG----- 941
Db 895 -----TEPIYQVAKKVKAKKIDRLDOIASGLGVGVQAGSFLLKRRHDKVDLDSKV 943
Qy 942 -----PLPKRHKVDLDSKVGLSRNQELQKIDNLNOAVSEAKAGF 982
Db 944 GLSANHEPIYATIDDLGGPPLKRRHKVDLDSKVGLSRNQELQKIDNLNOAVSEAKASH 1003
Qy 983 FGNLEQITDKLSDTKHNMPLWVESAKKVPASLSAKLDNYATNSHTRINSNKGAIN 1042
Db 1004 FDNLDQMDLKDSTKKNVNLVYVESAKKVPYLSAKLDNYATNSHTRINSNKGAIN 1063
Qy 1043 KATGMLTKQNPWKLVNDKIVAINVSGVPSLEYDKIGFNOKNMKDYSDSKFKSTKLNA 1102
Db 1064 KATGMLTKQNPWKLVNDKIVAINVSGVPSLEYDKIGFNOKNMKDYSDSKFKSTRLSNA 1123
Qy 1103 VKDNTSGFTQTLNFASTASYCIARENAEHGKIKNVNTKGGFQKS 1147
Db 1124 VKDIKSGFVQLTNIFSMGS-YSLMKASVEHGVRNTKGGFQKS 1167

RESULT 5
USOL_YEAST
ID USOL_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INT1 OR YD1058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
```

```
[3]
RN RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54378; CAA38253.1; -
CC EMBL; L03188; AAB00143.1; -
CC EMBL; U53668; AAB66659.1; -
CC PIR; A38455; A38455.
CC HSP; P80220; IDIP.
CC SGD; S0002216; USOL.
CC InterPro; IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 5.48; Score 314; DB 1; Length 1790;
Best Local Similarity 18.68; Pred. No. 0.00032;
Matches 233; Conservative 212; Mismatches 475; Indels 330; Gaps 47;

Qy 3 NEDSIQDQQTAEAFNPQOFINNQLQVAFKVDN-----AVASVDPDQKPI-----VD 48
Db 677 NEDSILPELDETGLPKVYFSTY---FIQFNENIYRIRLTALSHDPDEPKISFEEVE 733
Qy 49 KNDNRNQAQEGISOLREYSNK-----ATKNPKKNQYFSDFINKNLKNLKNL 99
Db 734 KLQRCQTKLKEITSLOTETESTHENLTKLIALTNEHKLDEKYQILNSSHSL-KENF 792
Qy 100 IDVESSTKSFQKFGDQ--RYRIFTSWSVSHONDPSKINTSRNFNMENIIQ-----PPIL 151
Db 793 SILETELKNVRDSDLEMTQRLDVKENOTALLEYKSTHKQEDSIKLEKLETLIL 852
Qy 152 DDKEKAE--FLKSAKQSFAGIIGNIQRTDQKFMGVDFESLKERQEAENKGEPTGGDWLD 209
Db 853 SQKKAEDGINKMKDLPF-----LSREMQAVENCK-----884
Qy 210 IFLSFIFDKKQSSDVKEAINOEPVPHVQPDIAATTTDIIQGLPPEARDLLDGRNFSKFTL 269
Db 885 -----NLQKEKDKSNVNHQKETSLSKEDIAAKITEIKAINENLEEMKIQCNLSK---934
Qy 270 GDMEMLDVEGVADIDPNYKFNOLLIHNNALSSVLMSGSHNGIEPEKVSLLYCGNGPGGARH 329
Db 934 -----NLQKEKDKSNVNHQKETSLSKEDIAAKITEIKAINENLEEMKIQCNLSK---934
```

```

Db 935 -EKEHISKELV-----EYK-SRFQSHDNLVAKL-----TEKLSLAN----- 969
QY 330 DWNATVGYKDOQGNVATINVHMKNGLVIAGGEGKINPNPFYLYKEDQLTGSQALRS 389
Db 970 -----NYKDMAENESLIKAV-----EESKNESIQL----- 996
QY 390 QBEIONKIDFMEFLAONNAKLDNLSEKEKEKEKFEKTEKDFOKD---SKAYLDALGNDRIF 446
Db 997 -SNLQNKIDSM-----SOEKENFQIERGSTIEKNIEQLKTKTISDLEQTKEBI 1041
QY 447 VSKKOTKHSALITEFGNGDLSVTLKYGKKADKALDRKN-----VTLGSLKHDGVVVF 501
Db 1042 ISKSDSSKDEYESQIS-----LLKEKETATTANDENVNKISLTKTREETEELEAELAYK 1095
QY 502 DYSNFKYTNASKNPNKPGVTVNCVSHL---EYGFNKAIFNLPDLNNAITSFVRNLE- 557
Db 1096 NLKNELETKLETSEKALKEVENEHLEKEEIQLEKEATEKQQLNSL-----RANLES 1149
QY 558 -----DKLTGGLSPQBAKLIKDFLSSNKELVCK 587
Db 1150 LEKEHEDLAAQLKYEEOIANKEROYNEEISQNLDEITSTQOENESIK---KKNDLEGE 1206
QY 588 T-----LNENKAVADAKNTGYDEVKKAOKDLE-----KSLRKRREHLEKEVEK 630
Db 1207 VKAMKSTSEEQNLKKSEIDALNL-QIKELKKKNETNEASLLESKSVESETVKIKELQD 1265
QY 631 KLESKGNKNKMEAKQAANSOKDEIFALINKNEANDARAIAQAQNLKGIKRELS----- 684
Db 1266 EGNFEKEVESELEDKLKASEDKNSYLELOKESK-----IKEELDAKTTEL 1312
QY 685 -----DKLENVKNLKDOPKDFPNKNGKNKDFSKAEETLKALKGVKDLGINPENISKE 740
Db 1313 KIOLEKITNLKAKESESELSRLKTKTSEERKNAEEOLEKL-----NETQ 1359
QY 741 NLNAALNEFKNGKNDKSVTQAKSDLENSVKDVII-----NOKYTDKVDNLNAQVSAK 795
Db 1360 INQAFERKERLLNKGSTITQIYSEKINTLEDELIRIQNENELKAKEIDNTRSELEKVS 1419
QY 796 ATGDFSERVEQALADLKNFS-----KEQLAQAQKNESLNARKKSEIYQSVKNGVNG--- 846
Db 1420 LSND-ELLEEKONTIKSLQDEILSYKDKITRDNKLLSIEDRKNRDL-ESLKEQLRAAQE 1477
QY 847 --TLVNGLSQAEATLTKNFSDIK--KELNAKLGNFNNNNNGLNKNEPIYAKVKNKKAAG 902
Db 1478 SKAKVEELGLKLEES-SKEAELEKSEMMKLESTIESNETELKSSMETIRKSDEKLE 1536
QY 903 QA-ASLEEPI-----YAQVAKVNAKIDRLNQAISGLGVVGOAAGFPL-----KRHDVQ 951
Db 1537 QSKSABEDIKNLQHEKSDLSIRINESEKDIIEELKSLRIEAK-SGSELETYVKQELNNAQ 1595
QY 952 DLSKVGLSRNQEALQKIDNLNQAVSEAKAGFGNLEQ-----TIDKLDSTKHNP 1002
Db 1596 EKIRINAEENTVLKSKLEIDIELKQAEIKSQNEKEELTSLRLKELEQELDSTQ----- 1651
QY 1003 NLWVESAKVPASLSAKLDNATNSHIRINSNIKNGATNEKATGMLTOKNPEWLKLVN-- 1060
Db 1652 ---QKAKSEBERAEYRKFE-----QVEKSQDEKAMLETKYN---DLVNKE 1693
QY 1061 ---DKTVAHNVGVPLESDYKIGFNQKNMKDYSDSEKFTKLNNAVKO 1105
Db 1694 QAWKREDTVTKTTDSQORQEIEKLAKELDLNKAENSKLKEANEDKSEIDD 1743

RESULT 6
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00799.
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
GN RBP2.
OS Plasmodium vivax (strain Belen).

```

```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M88098; AAA29744.1; -.
KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.2%; Score 307; DB 1; Length 1251;
Best Local Similarity 19.0%; Pred. No. 0.00038;
Matches 255; Conservative 221; Mismatches 471; Indels 392; Gaps 65;

QY 7 DOQPTAEAFNPQPFINLQVAFKLVDAVASYDPDQKPIVDKNDNRDNRQAFEGISQRLR 66
Db 3 DKEKVDTSFDEKK--KSIEKAYEKMGNTLKE-----LEKMD-DEKNIEKEVEEAAQI 51
QY 67 EYSNKAINKPTKKNOYSDFFINKSNDLINKNLII-----DYESSTKSFQKFGD 114
Db 52 QY-----KRIFIDHVDNLMNDEVEKSIIVMEKIELYKEIDEIKOKTNEVKQGD 101
QY 115 QYRIFTSWVSHQNDPSKI-----NTRSTRNFMENIIQPP-----ILDDK 154
Db 102 SNFYTEQYNSATOSKAKIEQFINIATTKGTSDTSQDINELESKEEVHKNLQVLKQES 161
QY 155 EKAEPKLSAKSAGFAGIIIGNQIRT-----DQKFMGVFDESLK----- 191
Db 162 NSMEEMRKQILSMKDLLILNSETIAKELISNTQNALG-FRENATKLNKTDDELLQVAA 220
QY 192 --ERQAEKNGEPTGGDWLDIFLSPFDKQSSDYK--EAINQEPVPHVQPDIAITTTT 247
Db 221 MIEEAKAHKN-----IDIALE---DAQIDTEVSKIEQINRE-IMNKKDEIKSYLSEI 269
QY 248 QGLPPEARDLLDERCNFSKFTLGDMEMLDVEGVADIDNYKFNOLLIIINNALSSVLMGSH 307
Db 270 K-----EYKDKCTTEISNKRCKDKTIEFLE----KFKFNEESNSKNVINEINIRSE 320
QY 308 NGIEPEKVSLLYGGNGPGARHDWNAITGVYKDOQGNVATINVHMKNGLVIAGGK 367
Db 321 QYLK-----DIEDAEKQASTKVELPHKH-----ETT 346
QY 368 INNPFYLYKEDQLTGSQALSQEIQNKIDFMELQAONNAKLDNLSEKEKEKFEKTEIKD 427
Db 347 ISN-----IFKESEILGVE-TKSQKKINKAEDIMKEIERHSEI-----QTQVK 390
QY 428 FQKDSKAY-----LDALGNDR-----IAFVSKKDTKHSALITEFGN-----GDL 468
Db 391 FOENLNKLNPHNYDNAEDLNNDKSTNAKVLITETNESVKHN--LSBITIKQGEKIY 448
QY 469 T-LKDYGGKADKALDRKNVTLOGSLKHDGVMFVDYSNFKVTN-----ASNPNKGVG-- 520
Db 449 SKANDIMQKIKATSENTAEKLE-KVKDDOSVYVNYLNQITERNLIYTERNLNGIDIST 507
QY 521 VTN-----GVSHLEVGFNKVAIFNLPDLNNAITSPVRNLEDKLTTKGLSPOEANKL 573

```

Db 508 ITNIEGALKESGNYEIGF-----LEKLEIG-----KNRKLKVDITKSI-----NST 551
Qy 574 IKDFLSSNKLVLGTLNFKAVADAKNTGNYDEVKKAQKLEKSLRK-REHLEKEVEKKL 632
Db 552 VGNESLFFNFDLNOYDFNKNINDYEN-----KMGIEYNEPEGSINKISELNASENTS 606
Qy 633 ESKSGNKNKMEAKAQAQNSQKDEIFALINK---EANDARAIAYAQNLKGIKRELSDKLENV 690
Db 607 DYNSAKTLRLLEA-----OKEV-NLLNKEEANKYLRDVKVESFRFI-FNMKESLDKI 658
Qy 691 NKNLKFDFKSPDEFKNGKNKDFSKAEETLKALKGSKVDGLGINPEWISKVENLNAALNEFK 750
Db 659 NEMIK-----KEQLTVNECHGNVOLVENIKELVDENLSDILKQ-A 699
Qy 751 NGKNKDFSKVTQAKSDLENSVDVI-----INQKVD----- 782
Db 700 TGKNEEQIKITH--STLKNKAKTILGHVDTSAKYVGKIKITPELATTELLGDAKLKTAQEL 757
Qy 783 KVDNLNOAV---SVAKATGDFSRVEQALADLNKFSKEOLAQO-----AOKNESLNARKK 833
Db 758 KFSKNNVVLTEMNKSNKTNELD-VHKNIQDAYKVALEILAHSDIEDTKDKDSSKLIEMG 816
Qy 834 SEIQSV-----KNGVNGTLVGNLSQAEATILSKNFSDIKK-----ELN 873
Db 817 NQIYLVVLINQYKNKISSIKSEAVSVKICNVSKKHSLSKITCSDKSYDNIIALEKQ 876
Qy 874 AKLGNFNN-----NNNN-----GLKN-EPIYAKVNNKKAG-----QAASL 907
Db 877 TELQNLNRSFTQKNTNTSDSKLEIKITDFESLKNALKTLGEVNAKASSDNHEHVQSK 936
Qy 908 EPIYAQVA---KKVNAKIDRLNQIASGLVGVQQAQFPFLKRDHVDLKLKSLGRNQELA 965
Db 937 SEPVPALSETEKEETDIDSLNTALDELLKKGRTG--EVSRYKLTIDVTREISDDTEL- 993
Qy 966 QKIDNLNOAVSEAKAGPFGNLEOTIDKL-----KDSKHPMNLWESAKKVPASL 1016
Db 994 --INTIEKNVAYLAYIKKNTEDVQDVLTLNEHFNKQVSNHEPTNF--DKSNK-----S 1045
Qy 1017 SAKLDNYATNSH-----IRNSNKNKAINKATGMLTOKNPEWLKLVNDKIVAH 1066
Db 1046 SEELTKAVTDSKTIISKLGVIIEVNEENTMTIESSAKEIEALYN-----ELKNKKTSLN 1101
Qy 1067 NV-----GSVPLESYDKIGFNOKNMKDYSDSPKSTKLNAVKDTNSGPTQLTNAFSTAS 1122
Db 1102 EYIQTSEVNEKLQE-----MKSNAKDYIDVKIENFTVLDTOK--SNIVTN----- 1143
Qy 1123 YYCLARENAEHGKINVTNK 1141
Db 1144 -----QHSINNVDK 1153

RESULT 7

YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 222.8 KDA PROTEIN CLF3.06C IN CHROMOSOME I.
GN SPACIF3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN {1}
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 270690; CAA94624.1; --
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 5.1%; Score 301; DB 1; Length 1957;
Best Local Similarity 18.3%; Pred. No. 0.001;
Matches 264; Conservative 240; Mismatches 489; Indels 452; Gaps 65;

Qy 2 TNETIDQO-POTEAANP-QQFINNLOVAFKVDNAVASYDDQKPIVDKNDNRNQAFE 59
Db 379 SNRTIHSQTLDAESKLSFSEQENKSLKGSIDYQNNLSSKDKMKVQVSSQLEEARSLAH 438
Qy 60 GISOLRE-----EYSNKAIKNPTKKNQYFDFIN-KSNDLINKNLIDVESTKSFQKFG 113
Db 439 ATGKLAEINSRDFQNKIKDFEIQDLRACLSSSNSELKEKSALIDKK----- 488
Qy 114 DQYRIFTSWYSHONDPSKINTRSIRNPMENIIOPIILLDDKEKAEFLKS----- 162
Db 489 DQELNNLREQIKQKKVSESTQSSLSQ-----LORDIILNKKKKHEVYESQLNELGELQ 543
Qy 163 -----AKQSPAGIIGNQIRTD-----QKFMGVFDE-----SLKREQAEKN 199
Db 544 EISNSEHSSLSQSLTLAAEKAATNNELSEKNSLQTLNFAOEKLAQSVOLKENSQN 603
Qy 200 GEPTGGWLDIFLSFIDFKOSSDVKEAINEPVPHPVDIATTTTIDIQLPPEAROLD 259
Db 604 -----FSSLDTSF-----KCLNESHQLENNH-----QTITKQLKDTSSKLQQL 644
Qy 260 ERGNF--SKFTLGD-----MEMLDV-----EGVADIDPN----- 286
Db 645 ERANFEQESTLSDENNDLRTKLLKLEESKSLIKQEDVDLSLEKNIQTLKEDLRKSEA 704
Qy 287 YKFNOLLTHN-NALSSVLGMSHNGIEPKVSLLYGGNGPGARHDWATVGYKDOGNV 345
Db 705 LRFKLEAKNLRVIDNLKKGKHELEAQ-----RNDLHSSLS----- 741
Qy 346 ATIIINVHMGSLVAGIEGKGINNPFFLYKEDQLTGSRALSQ-----EEIQN- 395
Db 742 -----DAKNTNAILSELTKSSEDDVKRLTANVETLTQDSKAMKQSFSLVNSYQISNL 795
Qy 396 ---KIDPMFLAQNNAKLDNLSEKEKEKFFTEIKDFOKSKAYLDALGNDRIAFVSKD 451
Db 796 YHELDDHVNMQSQNTLLS-----ESKLTDCENLTQNNMTLI-----DNVQKLMEKH 845
Qy 452 TKHSALITEFG--NGDLSYTLKDYKKADKAL-DREKNVTLQGSGLKHGVMFVDYSNFKY 508
Db 846 VNQESKVELKEVNGKSLDLKLRSSLNVAISNDQILTQLAELSK-----NIDSLEQ 899
Qy 509 TNASKNPN-KGVGVTVNGVSH-----LEVGFNKVA-----IFN 539
Db 900 ESAQLNSGLKSLEAKQLLHTENEELHRLDKLTGKLAIEESKSSDLGKLTARQEEISN 959
Qy 540 LPDLN---NLAITSFVRNLEDKITTKGLSPQEAANKLIKD-----FLS 579
Db 960 LKEENMSQSAITS-VKSKLDETLS-----KSKLEADIEHLKKNKVSSEVERNALLA 1011
Qy 580 SNKELVGKTLNFKAVAD-----AKN-----TGNYDEV-----KKAQKOLEK 616
Db 1012 SNEKMLDMLKNGENIASLQTEIEKKRAENDDLQSKLSVSEYENLLILLSSQNTKSLED 1071
Qy 617 SLRREHLEKEVEKKLESKKNKME-----AKAANSQ-KDEIFALINKEANDAR 668
Db 1072 KTNQLKYEKNVQKLLDEKQDNVLEBELTSKYGKLGGENAQIKDELLALRKKSKQHD 1131
Qy 669 AIAYAQNLIKGIKRELSDKLENVNNKLDKDFDKSFDEKFKNGKNKDFSKAETLALKGSVKD 728

```
Db 1132 CANFYDCLK-----EKSDALEQTNKNEKLVLSLEQSNNEALVEERSDLNRLSDMKKS 1187
QY 729 LGINPEWISKVEN-----LNAALNEFKGNK--DFSKVTOAKSDLENSVK-----772
Db 1188 LSDSNVISVIRSLRVNDELDTLKKDSDLSQYSEVCQDRDDLLSLKCEESFNKY 1247
QY 773 -----DVLIQKVTDK-----VNLNQAVSVAKAT 797
Db 1248 AVSLRELCTKSEIDVPVSEILDNFEVFNAGNFSLSRLTVLSLENYLDAFNQV-----1300
QY 798 GDFSRYE-----QALADLNKFSKEQLAQQAQKNSLNARKSEIYQSVKNQV 844
Db 1301 -NFKMELDNRLTTDAEFTKVADLEKIQHEHDDWLTQRGDLERALKDSE-----KNFL 1354
QY 845 NGLTVGNGLSQAET---TLKSNFSDIKKELNAKLGNNNN--NNGLKNE-----890
Db 1355 R-----KEAMTENIHSLEGKEETKETAELSRLLEDNQLATNKLKQDLHNLGEI 1406
QY 891 PIYAKVKKKQAGAAASLEPIYQVAKVNAKIDRLNQIASGLGVVGOAGPPLKRHDKV 950
Db 1407 RLKEDVLKEKESLIISLESLSNQ-RQKESLUDAKNEL-----EHLML 1448
QY 951 DDLKSVGLSRNQLAQKIDNLAQVSE-----AKAGFGNLE---QTIDKLKQST 997
Db 1449 DDTSR-----KNSLMKEKTESINSDDKSFELASAVEKLGALQKLHSESLSMENIKSQL 1504
QY 998 KINPNMLWESA--KKVPASLSAKLDNYA-----TNSHIR-TNSNIK--NGAINEKATG- 1046
Db 1505 QEAKKIQVDESTIQELDHEITATSKNNYEGKLNKDSDIIRLSENLQNLNLAEEKSAV 1564
QY 1047 --MLTOKNPEWLKLVNDKI-----VAHNVGSVPLSEYDKIGFN--1082
Db 1565 KRLSTEKESEILQF-NSRLADLEYHKQSEBELGRSLKSLASTTEELQALAEERLSLTR 1623
QY 1083 ----QKNMKOYSDSKFSTKLNNAVKDINS---GTFQPLTNAFTASYCYCLARENAHGI 1135
Db 1624 MLDLQNVKDLNS-----IKDSLSEDLRTLRLSLEDVSVASLQKECKIKSTNTVESL 1672
QY 1136 KNVNT 1140
Db 1673 QDVL 1677

RESULT 8
GOG4_HUMAN
AC Q13439; Q14436; Q13270; Q13654; PRT; 2230 AA.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY A 4 (TRANS-GOLGI P230) (256 KDA
GN GOLG4)
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
```

```
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
CC GOLGI.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE
CC SPLICING.
CC -!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
CC IN HEPATITIS B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U41740; AAC50434.1; -
CC EMBL; X82834; CAA58041.1; -
CC EMBL; U31906; AAC51791.1; -
CC EMBL; X76942; CAA54261.1; -
CC MIN; 602509; -
CC InterPro: IPR000237; GRIP.
CC Pfam: PF01465; GRIP; 1
KW Golgi stack; Antigen; Coiled coil; Alternative splicing.
FT DOMAIN 133 237
FT COILED COIL (POTENTIAL).
FT DOMAIN 276 1011
FT COILED COIL (POTENTIAL).
FT DOMAIN 1033 1214
FT COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152
FT VARSPLIC 2154 2185
FT TPYKGNLYHTDVSLEGFETFEYLKRVLEFY -> HLTKV
FT AICTIRMSHLENLPNLSICEKCFLSI (IN ISOFORM
FT 2).
FT MISSING (IN ISOFORM 2).
FT MISSING (IN ISOFORM 3).
FT FTSPRSKGF -> SWLRSSS (IN ISOFORM 4).
FT R -> K (IN REF. 3).
FT Y -> H (IN REF. 3).
FT T -> A (IN REF. 3).
FT K -> E (IN REF. 3).
FT T -> A (IN REF. 3).
FT K -> E (IN REF. 3).
FT K -> N (IN REF. 3).
SQ SEQUENCE 2230 AA; 261139 MW; 3BB73DB1EA86134 CRC64;

Query Match 5.0%; Score 291; DB 1; Length 2230;
Best Local Similarity 18.0%; Pred. No. 0.0026;
Matches 245; Conservative 249; Mismatches 487; Indels 380; Gaps 59;

Qy 4 ETIQDQPTQTEAFNPQPFINNVLVAFKVDNAVASYDPDQKPIVDKNDRDN---RQAFEG 60
Db 44 EQLDEGTPNRESGDTQSFPAKLQLRVPSVESLFRS--PIKESLFRSSKESLVRTSSRES 101
Qy 61 ISQLEEEYSNAIKNPTKKNYFSDFTINKSNDLINKNDLID----VESSTKSFQ-KFGDQ 115
Db 102 LNRLLDSSTASFPDPPSDMDSEADLVGNS-DSNLKLIQLRLRMKESLSYRGKYS- 159
Qy 116 RYRIFTSWVSHQNDPSKI-----NTRSRNFMENIQTPIIDDKKAEFLK-----161
Db 160 ---LVTAQOMLQREKKKLOGILSOSQDKSLRIAE--LREELQMDQQAKKHLQEEFDASL 214
Qy 162 SAKQSFAGI-----IIGNQIRTDQKFMGVFD-----ESLKERQEAENKGEPTGCD 206
```

Db 215 BEKQYISVLQTVSLKQRLNRGPMNVVLKPLPQLEPQAEVFTKEENPESDGEPPVED 274
Qy 207 WLDIFLSIFD---KKQSDVKEAINPVPVHPQPDIAITTTDQGLPPEARLLDERGN 263
Db 275 GTSVKTLETLOQRVKRQENLLKRC--KETIQSHREQCTLLTSEKAL-----QEQLDER-- 326
Qy 264 FSKFTLGDMEMLDVEGVADIDPNYK---FNOLLIHNNALSSVLMSGHNGIEPEKVSLLYG 320
Db 327 -----LQELEKIDLMHAEKTKLITQRLDAKNLIEO-----LEQDKGMV1-- 366
Qy 321 GNGPGARHDNATVGYKQDGNVATTIIVHMKNGSL-----VIAGEKKGINN 370
Db 367 ---AETKROMHETLEKKEEIAQLRSRIKQMTTQGEELREQEKESRAAFEELEKALST 422
Qy 371 P-----SYLYKEDQLTQSQRALSOEII-----ONKIDMEFLAQN--AKLD 411
Db 423 AQKTEEARRKLAEMDEQIKTIETSEBETISLQOELSRVQEVVDVNMKSSSEQIAKLQ 482
Qy 412 NLSEKE-----KEKPRTEIKDFORDSKAYLDALGNDRITAFVSKKDKTHS----- 455
Db 483 KLHEKELARKEQELTKLQTREREFQEQMKVALEKSOSEYLKISOEKBQESLALALEL 542
Qy 456 ---ALITFEGNDLSYTLKDYKGRADKALDRKNVTLOGSLKHGDGMVFDVSNFKYTNAS 512
Db 543 QKAILTESEN-----KLRLDQEAETV--RTRILELESSLE-----KSLOEN 583
Qy 513 KPNKGVGTNGVSHLEVGFNKVAIFNLPLNLNLAITSFVR-----RNLEDKLTFK 563
Db 584 KNQSKDLAV-----HLEAEKKN-----HNKEITVMVEKHKTELESKHQODALWTE 629
Qy 564 GLS-----POEANKLIKD-----FLSNKELVGKTLNPNKAVADAKNTG 602
Db 630 KLQVLKQOYQTEMKLEKREKCEQETLLKQKEIIFQAHIEEMNEKTLB-----KLDVKQT- 684
Qy 603 NYDEVKAAQKLSLRKREHLEVEVEKLESKGNKNKMEAK--AOANSOKDEIFALIN 660
Db 685 ---ELESLSSELSVILKARHKLEELSVLKQOTDKMKQLEAKMDEQKNHHQOQVDSIIK 741
Qy 661 KEANRDARAIYAQNLKGIKRELSKDLNVNKNLKDDEKS-----700
Db 742 E-----HEVSIQRTKALKADQINOLELLKRLKDKHLEKHAQVHENLEADIKRSEG 791
Qy 701 -----FDEFKNGKNKDFSKAETLALKGSKVKDLGI-----731
Db 792 ELQQAASAKLDVFSQSAETHQTKAYESQLAQLOOQLLDLETERILLTKQVAEVEAOKKD 851
Qy 732 -----NPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLEN 769
Db 852 VCTELDAHKIQVDLMQOLEKQNSMEMQKVKSLTQVYESKLEDGKGEQTKQILVKEEN 911
Qy 770 SV-----KDV-IIINQKVTDKVDN---LNOAVSVAKATGDFSRVEQALADLKNFSKEQ 817
Db 912 MILQREGOKKEIEILTQKLSAKEDSIHLNEEVET-----KFRNQEKMKVKQKAKEM 966
Qy 818 LAQQAQKNESLN--ARKSEIYQSVKNGVNTLVNGLSQAEBATLSKNFSIDIKKELNAK 875
Db 967 --QETLKKLLDQEAELKEL-----ENTAL--ELSQEKQKQFNAK 1002
Qy 876 LGFNENNNNGLKNEPIYAKVKKKAGQAAASLEPIYAQVAKKVNAKIDRLNQIASGLGV 935
Db 1003 MLEHAQANSAGISDAVSRLTNQKE--QIESLTVHRRELNDVISTWEKLNQQA----- 1055
Qy 936 VQGAAGPPLKRHKVDLKSGLVLRNQ-----ELAQKI-----DNLNOAVSEAKAGPFG 984
Db 1056 -----EELQETHIQLQKEQEVAELKQKILLFCQEKEMKKEITWLKEGV- 1102
Qy 985 NLEQTDIKLKDSTHNPMLNVESAKKVPASLSAKLDNYAT--NSHTRININNGAINE 1042
Db 1103 KQDITLNLQBLQKOKSAH--VNSLAQDETTLKAHLEKLEVDLKNLSKENTFLQEQVL 1160
Qy 1043 KATGMLTOKNPWLKLVNDKIVAHNVGSRPL--SEYDKIGFNKNKMDYSDSFK-----FS 1096
Db 1161 K---MLAEDRKRVSELTSLKLTDEEPQSLKSSHEK---SNKSLDKSLFKKLSSELA 1214

Qy 1097 TKLNNVAKDTNS---GFTQFLTNAFSTASYCYCLAR-ENAEH 1133
Db 1215 IQLDICCKKTEALLEAKTNELINISSKTNAILSRISHCOH 1255

RESULT 9

CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.F., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15005; CAA78727.1; -
DR PIR: S28261; S28261.
DR HSP; P03069; 1211.
DR MIM: 117143; -
DR InterPro: IPR001752; kinesin.
DR Pfam: PF00225; kinesin; 1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

Db	1935	EHKETVDKLRKISEKTIQISDIQKDLKSKDELQKKIQELQKKELQLLRVKEDVNM	SHK	1994
Qy	900	KAGOAASLE---EPIYA-----QVAKKVNKADIRLNOIASGLGVGQAAGFPLKR	HD	948
Db	1995	KINEMEQLKQFENFYLUCKGEMDNFQUTKKLHESLEIRIVAK-----		2037
Qy	949	KVDDLKSVGLSRNOELAKIDNLNOAV-----SEAKAGFFGNLEQITDK		992
Db	2038	ERDELRIKESLKNERDQFATLRMIARDQRNOHVQPKRLLSGQOHLMESLREK	CSR	2097
Qy	993	LKDKTK--HNPMLNVESAKKVPASLSAKLD-----NYATNSHIRINSNIKNGA	INEK	1043
Db	2098	IKELLKRYSEMDHDHYECLNRLSLDLLEKEIEFHRIIMKKLVVSVYTKIKEE--	QHECINKP	2156
Qy	1044	ATGMLTKQNPWLKLYNDKIVAHNVGSPVSEYDKTGFGN-----KNMKDYS	DSDFKFS	1096
Db	2157	EMDFIDEVEKQEKLLIKIQLHQQDC-DVPSRELRDLKLNQMDLHIEILKDF	SESEFPS	2215
Qy	1097	TKLN-NAVKDTNSGTFQFLNFASTASYCYCLARENABHGKKNVTK		1141
Db	2216	IKTEFQVLSNRKEMTQFLEWLNTR-----FDIEKLKNGIQKENDR		2257

RESULT 10

MLP1 YEAST

ID

MLP1 YEAST

STANDARD;

PRT; 1875 AA.

AC

Q02455;

DT

01-OCT-1993 (Rel. 27, Created)

DT

01-JUN-1994 (Rel. 29, Last sequence update)

DT

20-AUG-2001 (Rel. 40, Last annotation update)

DE

MYOSIN-LIKE PROTEIN MLP1.

GN

MLP1 OR YK8095W OR YKR415

OS

Saccharomyces cerevisiae (Baker's yeast).

OC

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OX

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

NCBI_TaxID=4932;

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=S288C;

RA

MEDLINE=93247549; PubMed=8483450;

RA

Koelling R., Nguyen T., Chen E.Y., Botstein D.;

RA

"A new yeast gene with a myosin-like heptad repeat structure.";

RL

Mol. Gen. Genet. 237:359-369(1993).

RP

[2]

RP

SEQUENCE FROM N.A.

RX

MEDLINE=94205265; PubMed=8154186;

RA

Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,

RA

Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;

RT

"The complete sequence of a 15,820 bp segment of Saccharomycetes

RT

cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three

RT

new open reading frames.";

RL

Yeast 9:1349-1354 (1993).

CC

!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA

CC

REPAIR.

CC

!- SIMILARITY: SOME, TO THE TPR ONCOGENE.

CC

!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".

CC

THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC

the European Bioinformatics Institute. There are no restrictions on its

CC

use by non-profit institutions as long as its content is in no way

CC

modified and this statement is not removed. Usage by and for commercial

CC

entities requires a license agreement (See <http://www.isb-sib.ch/announcement>

CC

or send an email to license@isb-sib.ch).

CC

DR

EMBL; L01992; AAA34783.1; -

DR

EMBL; X73541; CAA51948.1; -

DR

EMBL; 228320; CAA82174.1; -

DR

PIR; S38173; S38173.

DR

SGD; S0001803; MLP1.

DR

Coiled coil; DNA repair.

FT

DOMAIN 69 487

FT

COILED COIL (POTENTIAL).

ET DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 4.8%; Score 284; DB 1; Length 1875;
Best Local Similarity 18.5%; Pred. No. 0.0037;
Matches 253; Conservative 225; Mismatches 493; Indels 392; Gaps 57;

QY 3 NETIDQOPOTAAAFNPQOFI-----NNLOVAFKVDNNAVASYDDQKPIVDKNDK 52
DB 375 NENLSAKSSDFELKKOLIKERTKEHLQNOIETFIVELEHKV-----PIIN---- 422
QY 53 DNQRAF-EGISQLREEYSNKAI---KNPTKKNOYFSDFINKSNLDINKNDLIDVESSTKS 108
DB 423 ---SFKERTDMLNENLNAALLLEHTSNEKNKAKVKNLNAKQKLVCEAN--DLQTLTKQ 476
QY 109 FOKFGDQ-RYRIFTSWSH-----QNDPSKINTRSIRNF 141
DB 477 RDLQCRQIYLLITNSVSDSKGPLRKEEIOFIQIMOEDDSTITESDQKVVTERLVEF 536
QY 142 MENIOPPILDDKEAEFLKSAKOSFAGILIGNOIIRTDQKFMGVFDESLSKEROE----- 196
DB 537 -KNIIQ---LQEK-NAELLKVVNR-MLADKLESKEKSKQSLQKIESETVNNAKAIITLK 590
QY 197 -EKNSEPTGDMWDLIFLFDKQSSDVKEAINQEPVPHVOPDIATTTTIOGLPPEAR 255
DB 591 SEK-----MDLESRIEELQLEBELKTSVPNEDASYSNVITIKQLTETKRDLESQVQ 641
QY 256 DLLDERGNSFKFTIGDMEMLDVEGVADIDPNYKFNOLLIHNNALSSVLMGSHNGIEPEKV 315
DB 642 DLQTRISQITRESTENMSLLNKE-IQDLYDSKSDISIKLGRKESKRILAEERFKLLSNLT 700
QY 316 SLLYGNGCGCARHDWNAVTVGKDOGNVATIIIVHKNKSGSLVIAGGKGIN---NPS 372
DB 701 DLTRAENDQLKRFDYLONTILK--QDSKTHETLNEYVSCKSLSIVETEL-LNLKEBQK 757
QY 373 FYLKEDQLTGSRALSQEE-----IQNKIDFMEFLAQ 405
DB 758 LRVLHKLKQLNKLSPKSDSLRIMVTLQTLQKEREDELLLETRKSCQKIDLE----- 813
QY 406 NNAKLDNLSEKEKEFRTE--IKDFQKDSKA-----YLDALGNDRIFAFVSKDKTKHSA 456
DB 814 -----DALSELKETSQKDHKIKOLEEDNNNIEWYQNKIEALKKDYESTVSDSKOTD 868
QY 457 LITEFGNDLSYTLKDYCK--KADKALDRKNV-----TLQSLKHGDMFVD-YSN 505
DB 869 I-----EKLOYKVKLSLEKEIEEDKIRLHTYVMDDETINDSLRKEKSKINLTDAVSQ 922
QY 506 FK-YTNASKNPNKGVTVNGVSHLEVGFNKVA--IFNLPDLNLAITSFVRNLEDK--- 559
DB 923 IKEYKDYETTSQSLOQTN--SKLDESFKFTNQIKNLT-----EKTSLEDKISL 971
QY 560 -----LTTKGLSQEAN--KLKDFLSSNKELVGTILNFNKAIVADAKN----- 600
DB 972 LKEQMFNLNLDLQKGMEXEKADFKKRISILQNNKNEVEAVKSEYKSLKIQNDLQD 1031
QY 601 -----TGNVDEVKKAQKLEKSLRK-----REHLEKEVEKLES 634
DB 1032 QTIYANTQANNYEQLQKHADVSKTISELRQLHTYKQVQVTLNLSRQLENALKENEKS 1091
QY 635 KSGNKNK-MAKAQANSQK-----EIFALINKAN----- 664
DB 1092 WSSQKESLLEQLDLSNRIEDLSQNKLLYDQIQIYTAADKEVNNTNGPGLNLTITLR 1151
QY 665 -----RDARAIYAQNLKGIK----- 680
DB 1152 RERDILTQVTVARDAKMLRQKTSIMDLVDLQDARTKLDNSRVEKENHSSIIQOHHIME 1211
QY 681 -----RE-----LSDKLENNKLNKDFDKSDFDEKKNKDFSKAEETLKALKGSKVD 728
DB 1212 KLNQNLNLSRITURNLENNNNKKKQLQSELDK-----QNVAPISELTALKYSMOE 1267

QY 729 LGINPEWLSKYVENLNAAALNEFKNGKNKDFSKVTOAKSOLSNESVNDVVIINQKVTDKVDNL- 787
DB 1268 -----KEQELKLAKEEVHRNKKRS-QDILEKHEQLSSS-----DYKLESEIENLK 1312
QY 788 ---NQAVSVAKATGDFSRVEQALADLKNFSKEQLAQAQAKNESLNARKKSEIYQSVKNG 843
DB 1313 ELENKQKQGAEEKFNRLRRQ-----AQRKLTSKLSQSLTEQVNSLRDA 1360
QY 844 VNGTLVGNLGSQAAATLTSKFNFSIDIKELNAKLNKFNFNNNNNGLKNEPIYAKVNNKKKQAG 903
DB 1361 KN--VLENSLSEANAR-----IEELONAKVAGNQNQLEAIRKLOEDAESKRE---L 1407
QY 904 AASLEEPI--YAQVAKVNAKIDRLNQIASGLVGVVQAAGFPFLKRDHVDLDSKVLGRN 961
DB 1408 QAKLEESTSYESTINGLNEEITTLKEIE-----KQRIQQOQLQATSANEQ 1454
QY 962 QELAOKIDNLQAVSEAKAGFFGNLEQTI-DKLKDKSTK--HNPMLWVESAKKVPAS--- 1015
DB 1455 NDLNIVESMKKSFEEKIKFIKEKTOEVNEKILEAQERLQNOPSNINNEIKKKWSEHE 1514
QY 1016 --LSAKL--DNYATNSHIRINSNIKNGAINKATGMLTKPKNPEWKLVDNKIVAHNVGSV 1071
DB 1515 QEVQKIREAEALKKRLPT-----EKKINKIIEKKEELEKEFEKVEERIKSME 1567
QY 1072 PLSEYD-----KIGFNQKNMKDYSDSPKFSFKLNNAVKD 1105
DB 1568 QSGETIDVVRKQLEAKVQEKOLEN-----EYNNKKLOEELKD 1605

RESULT 11
YIO9_YEAST
ID YIO9_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
GN YIIL149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Consey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DSC-1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z38059; CAA86129.1; -
DR PIR; S48385; S48385.
DR SGD; S0001411; MLP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 4.8%; Score 279; DB 1; Length 1679;
Best Local Similarity 18.5%; Pred. No. 0.0049;
Matches 258; Conservative 222; Mismatches 490; Indels 426; Gaps 56;

QY 19 QOFINNVAFLKVDNAVASYDPDQPIVDKNDNRNDAFEGIS-----OLREE----- 67
 Db 38 EBEVTKNLVYDVEIKSQYYSRISKQLQDLDESSEQNTAKEELGKDLQOLNEERSYRRE 97
 QY 68 -----YSNKAIK--NPTKK-----NOYFSDFTKNSDLIN----- 95
 Db 98 IDALKQLHVSHEAMREVDKRVKEEYDIWQSRQOQNDLNDLKNKLLRRKLMEME 157
 QY 96 -----KONLID-----VESSTKSFQK-----FG 113
 Db 158 NILQCKSNALSQLKYDTVSQEKELMLQSKLIEBKLSFSKTKLTBEVTKSSHVENLE 217
 QY 114 DORYRI-----FTSWSHONDPKINTRISINFMENIIPILDDK-----EKAFLK 161
 Db 218 EKLYQMOSNYESVFTYNKFLNQN---KOLQSQVE---EKVLEMKNLKDTASVEKAERFS 271
 QY 162 S-AKOSFAGIIGNOIRDTQPMGVFDESLEKQEAENG--EPTGGDWLDFISFIDK 218
 Db 272 EMTLOKNNDLLRSOLTSEK-----DCSLRAIEKNDNDSNCRNPEHTVIDELIDTKURL 326
 QY 219 KOSSD-----VKEAINQEPVPHVQPDIAATTTDIIQGLPPEARDLLDERGNFSK 266
 Db 327 EKSNECORLQNVMDCTKEBEATWTSVAVSPTVGKLFSDIKVL---KROLIKERN--QK 381
 QY 267 FTLG-----DMEMDVEGVADIDPNYKFNOLLIHNNALSSVLMGSHNGIEPEKVSLL 318
 Db 382 FOLQNOLEDIFILEHKTPELISFKERTKSLHEHLKRSTELLETVSLTKRQEREITSRL 441
 QY 319 YGNGGPGARHDWATVGYKQOQGNVATIIIVHMKNGSGLVIAGGEKGINPSFYLYKE 378
 Db 442 QKING-----CEANIISLVKQRLDLARQVKLLNLSAI-----QETASP-----LSQ 484
 QY 379 DOLTSORALSQOEIQONIKDFMEFLAONNAKLDNLSEKKEKFRPEIKDFQDKSKAYLDA 438
 Db 485 DELISLRKILESNNVENDSQAIIITERLVEFSNVNELQERN-----VEL 529
 QY 439 LGNDRIAFVSKDKTHSALITFEFGNDLSYTLKDYGGKADKALDREKNVTLOGSLKHGV 498
 Db 530 LNCIRI-----LADKLENYEGQDKTLQKVENQTIKEA--KDAI 566
 QY 499 MEVDYSNFKYTN-----ASKNPNGVGTNGVSHLEVGFNK-----VAIF 538
 Db 567 TELENINAKMETRINILLRERDSYKLLASTEENK--ANTNSVTSMEAREKIRELEAEL 624
 QY 539 NLPDLNLAITSFVR-----NLEDKLTITKGLSPOEANKL-----IKDFL 578
 Db 625 SSTKVENSAILONLRKELLYKKSOCKKTKTLEDFENFKGLAKEKMERLEAIDHILKAE 684
 QY 579 SSNKELVGKTLNFKAVADAKNTGNYDEVKKAQDKLESLRK-----REHL----- 624
 Db 685 EKQKSWPSPYIHVEKERASTELSQSRIKIKSLEVEISK-LKKETASFTPTRESLTRDPEQ 743
 QY 625 ----EKEVEKKLESK--SGNKNMEAKAQAQNSQDEIFALINK-----EA 663
 Db 744 CKEKKELOMRKESEIHNENKNDKFSQEGOYKAKIKELENNLERLSQLSQIKQIES 803
 QY 664 NRDAR--AIAYAQN-----LKGIKRELSDK---LBNVKNLKDFOKSFDEPK----- 705
 Db 804 IRSCKDSQLKWAQNTIDDTMKMSLLTSLNSKETTIEKLSSEIENLDKELRKTQYQV 863
 QY 706 NGKNKDFSKAETLK-----ALKGSVKOLGINPENWISKVENUNALNEFKN----- 751
 Db 864 LDONSDASTLEPTLRKELEQIQVOLKDKANSQIQAYEEIISNNEN---ALIELKRELAKTK 920
 QY 752 -----GKNKDFSKV-----TQAKSDLENSVRDVI 775
 Db 921 ENYDAKIELEKKEKWARDLEDLSRLGELGETRALQPKLKEGALHFVQOSEKRLNEVERKI- 979
 QY 776 INOKVTDKVDNLNOAVSAKATGDFSRVEQALADLNKFSK--EQLAQQAQKNESLNARKK 833
 Db 980 --QKMIKIEKMTIVQLCKKK-EMSOYQSTMKENKNDLSSELVIRLEKDAADCAQELTKTK 1036
 QY 834 SEIQSVKNGVNGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNFNNNNNNGLKNEPIY 893

Db 1037 SLYSA-----QDLIDKXHERKWEKADYEREL---ISNIEOTESURVENSILI 1082
 QY 894 AKVNKKKA--GOAASLE-EPIYAQVAKVNAKIDRLNQLIASGLGVVGOAAGPLKRHDKV 950
 Db 1083 EKVDDTAANGDKOHLKLVLSFSLNHRNSLEYTKLTCTCKRELAFVKQ-----KN 1132
 QY 951 DLSKV--GLSRNQLAQKIDNLNOAVSEAKAGFFGNLEQITDKLKOSTKINPMNLWVES 1008
 Db 1133 DSLEKTINDLQRTQTLSEKQCSAVI-----IDEPKDTIK----- 1168
 QY 1009 AKVPASLSAKLDNATNSHTRINSNIKANGAINKATGMLTKQKNPEWKLKLVNDKIVAHNV 1068
 Db 1169 -----EVTQVNILKENNAITLOKSLKNVTEKKNREIYKQLNDR--QBEI 1208
 QY 1069 GSVP---LSEYDKIGFNOKMKDYSDFKSTK-----LNNAVKDTNSGFTOFLTNAFSTA 1121
 Db 1209 SRLQDRLIQTQEQVINSNKILVYESEMEQCKQRYQDLSQOQKDAQKDKIEKLTNEISDL 1268
 QY 1122 SYCLARENAEHGKKN 1137
 Db 1269 KGKLSSAENANADLEN 1284
 RESULT 12
 ID RBPI_PLAVB STANDARD; PRT; 2869 AA.
 AC 000798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBPI.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M88097; AAA29743.1; -;
 DR HSP; P36956; IAW9;
 DR Malaria; Receptor; Signal; Transmembrane.
 KW SIGNAL
 FT CHAIN 1 17 POTENTIAL.
 FT DOMAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT TRANSMEM 18 2807 EXTRACELLULAR.
 FT DOMAIN 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOPLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 4.7% Score 276.5; DB 1; Length 2869;
 Best Local Similarity 19.1%; Pred. No. 0.011;
 Matches 265; Conservative 196; Mismatches 451; Indels 479; Gaps 63;

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X53947; CAA37894.1; -;
 EMBL; X06187; CAA29550.1; -;
 EMBL; U10399; AAB68872.1; -;
 PIR; S05806; S05806;
 PIR; S12323; S12323;
 PIR; S46773; S46773;
 HSP; P08799; IMND;
 SGD; S0001065; MYO1;
 InterPro; IPR000048; IQ;
 InterPro; IPR001609; myosin_head;
 Pfam; PF00063; myosin_head; 1;
 PRINTS; PR00193; MYOSINHEAVY;
 ProDom; PD000355; myosin_head; 1;
 SMART; SM00015; IQ; 1;
 SMART; SM00242; MYSC; 1;
 PROSITE; PS00096; IQ; 1;
 Myosin; Actin-binding; ATP-binding; Coiled coil; Alkylation.
 DOMAIN 1 793
 FT DOMAIN 794 823 IQ.
 FT DOMAIN 856 1911 COILED COIL (POTENTIAL).
 FT NP_BIND 180 187 ATP (BY SIMILARITY).
 FT DOMAIN 460 529 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 703 703 ALKYLATION (BY SIMILARITY).
 FT CONFLICT 36 36 K -> I (IN REF. 1).
 FT CONFLICT 46 46 I -> T (IN REF. 1 AND 3).
 FT CONFLICT 59 59 V -> S (IN REF. 1 AND 3).
 FT CONFLICT 86 86 L -> F (IN REF. 1).
 FT CONFLICT 330 330 MISSING (IN REF. 1 AND 3).
 FT CONFLICT 343 343 N -> S (IN REF. 1 AND 3).
 FT CONFLICT 421 426 QAKFI -> TKLSSL (IN REF. 1).
 FT CONFLICT 515 515 D -> S (IN REF. 1).
 FT CONFLICT 529 535 SKGPTG -> ARGHDR (IN REF. 1 AND 3).
 FT CONFLICT 541 541 D -> V (IN REF. 1 AND 3).
 FT CONFLICT 550 551 TD -> LM (IN REF. 1).
 FT CONFLICT 573 573 R -> A (IN REF. 1).
 FT CONFLICT 582 582 H -> D (IN REF. 1).
 FT CONFLICT 588 599 EYVEGWSKNNK -> NTLWAGVPKT (IN REF. 1).
 FT CONFLICT 599 599 MISSING (IN REF. 3).
 FT CONFLICT 627 632 EKSSA -> GKLLVC (IN REF. 1 AND 3).
 FT CONFLICT 695 695 R -> S (IN REF. 1 AND 3).
 FT CONFLICT 736 742 ENSTTT -> RKENHD (IN REF. 3).
 FT CONFLICT 756 756 E -> R (IN REF. 1 AND 3).
 FT CONFLICT 773 784 NTKLFFRAGVLA -> ILTVFKLEYWS (IN REF. 1).
 FT CONFLICT 793 794 KL -> NV (IN REF. 1).
 FT CONFLICT 896 896 N -> T (IN REF. 1).
 FT CONFLICT 900 900 N -> NSQITKINTITETPOSTVIGERPKRVICGN (IN REF. 1).
 FT CONFLICT 906 906 N -> I (IN REF. 1).
 FT CONFLICT 911 911 N -> K (IN REF. 1).
 FT CONFLICT 915 930 NESLNRVTSSETLQ -> RIAKILKPAINIT (IN REF. 1).
 FT CONFLICT 934 939 DDIVSE -> MTLFL (IN REF. 1).
 FT CONFLICT 951 953 AON -> RKI (IN REF. 1).
 FT CONFLICT 955 958 EEAH -> KKL (IN REF. 1).
 FT CONFLICT 1002 1002 S -> C (IN REF. 1).
 FT CONFLICT 1049 1049 L -> D (IN REF. 1).
 FT CONFLICT 1056 1056 C -> S (IN REF. 1).
 FT CONFLICT 1060 1060 M -> I (IN REF. 1).
 FT CONFLICT 1085 1085 A -> E (IN REF. 1).
 FT CONFLICT 1123 1123 V -> C (IN REF. 1).
 FT CONFLICT 1133 1133 L -> S (IN REF. 1).
 FT CONFLICT 1144 1146 KSN -> NLI (IN REF. 1).
 FT CONFLICT 1159 1168 RETKEQPKK -> TRKEQDKK (IN REF. 1).
 FT CONFLICT 1179 1181 SKI -> ELKV (IN REF. 1).
 FT CONFLICT 1184 1185 LE -> WK (IN REF. 1).
 FT CONFLICT 1188 1204 LSEISLNLQYLNKRISG -> CHRKYLSILKQKNIR (IN REF. 1).

FT CONFLICT 1224 1224 P -> S (IN REF. 1).
 FT CONFLICT 1228 1228 E -> Q (IN REF. 1).
 FT CONFLICT 1253 1253 E -> Q (IN REF. 1).
 FT CONFLICT 1311 1323 PKESDINKLWLE -> LTKSLIILTNAGS (IN REF. 1).
 FT CONFLICT 1400 1400 D -> H (IN REF. 1).
 FT CONFLICT 1454 1554 SEQDLRQKOLESTERQKELLSSTIKQKQOFENCMDLQ
 NELRLREHIALKQAEDVKNMASIIIEKLTQNKQKELI
 ERMERNDSQMLQETLE -> P (IN REF. 1).
 FT CONFLICT 1568 1568 D -> V (IN REF. 1).
 FT CONFLICT 1630 1646 DLLKQLDHYTKVVEMLN -> SEARSLYKSGGND (IN REF. 1).
 FT CONFLICT 1698 1704 MISSING (IN REF. 1).
 FT CONFLICT 1725 1737 TLQMQEONSNG -> NTTANGTKPKREW (IN REF. 1).
 FT CONFLICT 1754 1757 FDDE -> LMW (IN REF. 1).
 FT CONFLICT 1777 1777 D -> E (IN REF. 1).
 FT CONFLICT 1788 1788 R -> T (IN REF. 1).
 FT CONFLICT 1825 1825 S -> D (IN REF. 1).
 FT CONFLICT 1882 1882 S -> W (IN REF. 1).
 FT CONFLICT 1902 1904 FWK -> NSGKRLDADDL (IN REF. 1).
 SQ SEQUENCE 1928 AA; 223634 MW; 6F54C7611F43DC9F CRC64;
 Query Match 4.7%; Score 276; DB 1; Length 1928;
 Best Local Similarity 20.8%; Pred. No. 0.0073;
 Matches 292; Conservative 208; Mismatches 479; Indels 426; Gaps 75;
 Qy 16 FNPOQFTNNLOVAFKVDNVAASVDPDOOKPIVDKNDNRQAFEGISOLREYSNKAIKN 75
 Db 256 FNEHGMINGAHIEW-----YLEKSRIVHQSKERN--YHIFYQLSLDGLDSELKN 304
 Qy 76 PTKKNVYFDF-----INKSNDLKNLDLVESSTK-----SFQKQDQRYRTF---- 120
 Db 305 LRLKSRNVKDKILSNSQDIIIPGIN--DVENFKELLSALNIIIGFSK--DQIRWIFQVVA 360
 Qy 121 -----TSWVSHQNDPFSKINTR---SIRPFMENIIOPILDDKE----- 155
 Db 361 IILLIGNIEFVSDRAEQASFKNDVSAICSNLGVDEKDFQTAILRPSRKAGKEWVSQSKNS 420
 Qy 156 -KAERFL-----KSAKQSAGI--LIGNQIRTDQKF----- 182
 Db 421 QOAKFILNALSRLNIYERLFGYIVDMINKNLDHGSATLNYIGLLDIAGFEFENNSFEOLC 480
 Qy 183 MGVFDESILKERQAEKNGEPTGTDWLDIFLSPDKKSDVKCAINOEPVPHVQPDIAI 242
 Db 481 INYTNEXLQ-----PFNNHMFVLEQSEYKLENIQWDYIDYCK-DLQL 522
 Qy 243 TTTDIIQ--GLPPEARDLLDERG-----NFSKFT----- 268
 Db 523 TIDLIESKGPPTGVPLLDDEAVLPKSTDESFYKSLISTWDQNSKPKRSRLKNGFILKH 582
 Qy 269 -LGDWEMLDVEG--VADIDPNYKFNQLLIHNNALSSVLGSHNGIEPKVSLLYGGNGCP 325
 Db 583 YAGDVE-YTVEGWLSKNKP-----LNDNLLSLSSSONDI-----ISKLFQPEGEK 628
 Qy 326 GARHDWATVG-----YKDOQGNVATIINVMKNGSLGITVAGEK 366
 Db 629 SSSAGVEANISNOEVKKSARTSTFKTSSRHREQOITLLNQLASTH-PHFVRCIIPNVK 687
 Qy 367 GIN--NPSFYLYKEDQL--TGSQRA--LSQBEIQNKIDFMEF-----LAONNAKLNDLS 414
 Db 688 KVKTFNRRLIL---DQLRCNGVLEGIRLAREGYPNRIAFQEFFQRYRLYLPSNSTTTF 744
 Qy 415 EKEKEKETE---IKDFQKDSKAYLDALGNDRIFAVS-----KKDTHSALITEFG 462
 Db 745 SKLKASTRQNCFEFLTSLQDITKVY--KIGNTKLFFKAGVLADLEKQKDVKNLNMIRK-- 800
 Qy 463 NGDLSYTLKDY--GKKADKALDREKNVTLOGS-----LKHGDMVFMVDYSNFKYTNASK 513
 Db 801 ---LTATIRGTVRKEITYHLOKLKTRKTRVIGNTFRNLRYKEDPWFNFIKPLLTSS 857
 Qy 514 NPNKGVGTNGVSHLEVGFNKVAIFNLPDLNLAITSFVRNLEDKLTKGLSPQEAANKL 573

Db 858 ND-----MTFTKFNQINKLK-----NDLOMESKKKFLSEK-----NOKTVNEL 898
Qy 574 --IKDFLSSNNKELVGKTLNFNAKAVADAKNTGNYDEVKKAQKDL--EKSLLRKRHLE---- 625
Db 899 ENTODLLNOEKE-----NLKRNESLNVRKTS--ETLQKQFDDLVSEKDEISREKLEVAQN 953
Qy 626 -KEYEKLKESGNNKMEAKAO--ANSQKDEIFALINKENRDRARATAYQANLKGIRREL 683
Db 954 LEEAHQIQGLQETIRREATLEKLHKKNNELIKQIS--DLNCD--ISKEQSOSLIKES 1009
Qy 684 SDKLENVKNLKDQDFDKNGKNDKDFKAEETLKALKGVKDLGINPEWISKVENLN 743
Db 1010 KKLKLEIKRLKDVINSKEBEIKSFNDKLSSSE-----DLDI--KLVTLEKNCN 1057
Qy 744 AALNEFKN-----GKNKDFSKVTQA--KSDLENSVKDVIINQKVTDKVDN----- 786
Db 1058 IAMSRLQSLVTSNDSLRKSNENFKKKAALNNLKNKESELL--KMKKEIDNHKKELAT 1114
Qy 787 ----LNOAVSV--AKATGDFSRVEQALADLK-----NFSKEQLAQQAQKN-----E 826
Db 1115 FSKORDDAVSEHGKITAELEKTRIQLTREYKSNYQKIKKEYSNFQRETKEQEOKKRNLSVE 1174
Qy 827 SLN-----ARKKSEIYOSKNGVNGTLVGNGL--SOAEATTLKSNFSD-----IK 869
Db 1175 SLNDSKIKELEARLSQEI--SLNOYLKRIKSGNSVETNISSTRSTSYSDPLDKEDIK 1232
Qy 870 KELNAKLGNFNNNNNGLKNEPIYAKVKKK-----AGQAAASLEEPIYAKVAKK 918
Db 1233 KYDQLA--FTEITRN--LENE-----IEEKNLISRLRFTETRIASSFED-----QK 1278
Qy 919 VNAKIDRLNQIASLG---VVGQAAGPLKR--HDKVDDLSKVLGRNOELAQKID----- 969
Db 1279 IKAQMKKLLIQMDPSIPLDSILNEPLDNCPOKESDINKLMLEVY--LKRQLDIETRA 1337
Qy 970 --NLNOAVSEAKAGF-----PGNLEOTIDKLKSTKKNPM-----NLW 1005
Db 1338 HYDENALISALHFKRIQIGESSLSDDIYKLPFEASEERVKSLDEKLTMLPLKDRTNLP 1397
Qy 1006 VESAKKVPASLSA-----KLDNYATNSHIRINSNIKNGAINEKATGMLOTOKNPEWLK 1057
Db 1398 VGDIKRNDSISKYEERYKYLENYKLEITL-----NESNGKLSQILDLRQSKSKEAL-- 1452
Qy 1058 LVNDKIHAHNVGVSPLSE--YDKIGFNQNMKDYSDSPKSTFKLNAVYKDTNSGTQFILT 1116
Db 1453 -----LSEQLDRL-----QKDLSETERQKELLSTIKQKQKQFENCMDD 1491
Qy 1117 AFST---ASYCYCLARENAEHCINW 1138
Db 1492 LQGNELRLREHIALKQAEEDVRNN 1516

RESULT 14

ANT1_ONCVO STANDARD: PRT: 2022 AA.
AC P21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MAJOR ANTIGEN.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; PubMed=7770081;
RA Tritreaprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA Neubert T.A., Scott A.L.;
RT "Molecular cloning of a gene expressed during early embryonic
development in Onchocerca volvulus.";
RL Mol. Biochem. Parasitol. 69:161-171(1995).

12]
RN SEQUENCE OF 733-866 FROM N.A.
RP MEDLINE=89127417; PubMed=2464764;
RA Luciferson J.E., Duke B.O.L., Moser D., Zeng W., Erundu N.E.,
RA Donelson R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
characterization of the cDNA for a major antigen.";
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U12681; AAA00009.1; -;
DR EMBL; J03995; AAA29412.1; -;
DR PIR; A54513; A54513.
DR HSP; P02633; 3ICB.
KW Antigen; Coiled coil.
FT DOMAIN 74 120 COILED COIL (POTENTIAL).
FT DOMAIN 151 251 COILED COIL (POTENTIAL).
FT DOMAIN 327 384 COILED COIL (POTENTIAL).
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 4.6%; Score 272.5; DB 1; Length 2022;
Best local Similarity 19.0%; Pred. No. 0.01;
Matches 225; Conservative 213; Mismatches 416; Indels 329; Gaps 57;

Qy 1 MTNETI-----DQOPTEAAFPNQFINNQVAFK-----VDNAVASVYDPQKPIV 47
Db 285 ITQDTVLIEAVKRPHESSQQTAPVIGP-ELITELRAEDAGLHDELMRYESAKRII 343
Qy 48 DKNDNRQAFEGISQLEESYNNK--AIKNPTKKNQYFSDFINKSNLDINKNLDIVESS 105
Db 344 ELEARD-----DESHNKLVALESCLKRTR--DRLAESQNALRK--LYDM-- 383
Qy 106 TKSQKQFGDQRYRFTSWSHONDPSKINTRSIRNF-----ENITQPTLLDDEKAEF 159
Db 384 TYSYEINAEKRETSPTKGYVPPPEVV--RSRYVLNSRANDNNVLQKLNKAEVQISE 441
Qy 160 LKSAQKQFAGIIGNQIRTDQKPMGVFDESUKQBAEKNGEPTGGDWLDFLSFIDPK 219
Db 442 LTTKNDLSEI-----RRLEKRIAEANRTITHQRELDLDAKHTVKD--LEDRLKSLQEK 495
Qy 220 QSSDVKEAINQEPVPHVQPDIAITTTTIDQGLPPPEAR-----LLDERGNFSKFTLGM-- 272
Db 496 ASIDSARHLEDEIRKMRQEPNSTLLDVERAAEDADERIRKIDEE--TKIRISELTNR 552
Qy 273 -EMLDVEG-----VADIDPNYK--FNOLIHNNALSSVLMSGH--NGIEPE 313
Db 553 IEMLLEENKRLKDENDGMKNRIQIDIEKYNTIIRKLEEKONALKNLENTROLVNELEE 612
Qy 314 KVSLLYGGNGPGARHDM-----NATVGYKDOGNVATTINV-----HMKNG 356
Db 613 R-----TRFDTMTSEFDNLTNTNDSANKNTVAIELTVKQKEQDEIKKDK 660
Qy 357 SGLVIAGEKGINNPSFYLYKEDQITQSQALSQEEIQNKIDFMFEFLAQNNAKLDNIS-- 414
Db 661 LAKELADLENKLNN-----ETKMRGDAEKLQNRHL--DEID--NFKQINYEIVETII 710
Qy 415 EKEKEKFEITEIKDQKSKAYLDALGNDRIFAIVSKKDKTKHKSALITEFGNGDLSYTLKDYG 474
Db 711 RRQNDDEFTQM---KTNQAKLSMKNSLIA--AKKETEK---LSEMNN----- 750
Qy 475 KKADKALDREKNVTLOGLKHGDMFVDYSNFKYTNASKNPKNGVGTNGVSHLEVGFNK 534
Db 751 -----RLQODKNDLIGAKQKQDTELNL-----LTEKIRKVEIEFER 786

Qy 535 VAINLPDNLNLAITSFVRRNLED-KUTTKGLSQEANK---LIKDFLSSNKELVGTLN 590
Dy 787 IKDN-----QLEDHERTARDLQKQETNRNHLAKLEEARADIV--ALN 830
Qy 591 FNKAVADAKNTGNYDE-VKKAQKOLEKSLRREHLEKEVEKKLESKGNKKNKAQ-A 648
Dy 831 DRLAKMDANFKIKLDETITKSPADHE-TIKSRESKSEIKVHKETKIYEINKYRAELEKL 889
Qy 649 NSQKDE----TFALINKEANDARAIAYAQNLKGIKIKLSLSDKLENVKNLKDFOKSF--- 701
Dy 890 ESKDDLEKRIIGLODELNEKDRDTRLNABIDDLKRLQTEIEKVRKETTVOERHIE 949
Qy 702 -DEFKNGKNDKFSKAEETLKALKSVKDL-----GINPEWISKVENLNA 745
Dy 950 WDEERDHNQKIDSMNALIDELRSKLNDRAERAMADLQNRDSILERENNWNKESDALNME 1009
Qy 746 LNEFKN---GNKDFSK-VTOAKSDLENSVKDVI-----INQV-----TDKVDNLQ 789
Dy 1010 LDRLELLSVRRDAEKEINRYNTDLQTAARNEIKLTPPTNMKSQLNAAEDKINSLNK 1069
Qy 790 AVS-----VAKATGDSRVEQALADLK-----NFSKEOLAQAQAKNESLNARKKS 834
Dy 1070 VITQONKIRDLTGEVHLELGDKAGNVANLESELDTTTHERIHLGEOQNASLQ----- 1124
Qy 835 EIYOSVKNVNGTGLVNGLSQAEATTLISKNFSDIKELNAKLGNNNNNNGLK--NEPI 892
Dy 1125 -----TELNK-----IKGDIDSLFG-----ENDMLKTAKESN 1151
Qy 893 YAKVNNKKAGQAASLEBPIYAQVAKVNAKIDRLNQIATSLGVVGAAGPLKRHKVD 952
Dy 1152 EAEIDRLKQKLOQRIEN-----AKKYSDALDKLR-----PEYDRLQ 1188
Qy 953 LSKVGLSRNQELAKIDNLNOAVSEAKAGPFGNLEQITDKLKDST-----KHNPMLN 1004
Dy 1189 LYR-----EKIKQAEMLTQAVODLES-----RLNQSRRELURDQTDKLIASEGDRNALRS 1237
Qy 1005 WESAKK-----VPASLSAKLDNY--ATNSHIRINSNIKNGAIN 1041
Dy 1238 EVEKLQHEVQPMREQLLRKTDQYQAALSDLVNAHRTAEDGRVN 1280

RESULT 15
PCPI_SCHPO STANDARD: PRT: 1208 AA.
AC Q92351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPINDLE POLE BODY PROTEIN PCPI.
GN PCPI OR SPAC6G9.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Flory M.R., Morphey M., Joseph J.D., Means A.R., Davis T.N.;
RT "Pcpip, a Spcl10p-related calmodulin target at the centrosome of the
RT fission yeast Schizosaccharomycetes pombe."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Connor R.E.;
RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPINDLE POLE BODY COMPONENT THAT BINDS CALMODULIN.
CC OVEREXPRESSION OF PCPIP CAUSES THE FORMATION OF SUPERNUMERARY SPB-
CC LIKE STRUCTURES AND DISRUPTS BOTH MITOTIC SPINDLE ASSEMBLY AND
CC CHROMOSOME SEGREGATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z81317; CAB03608.1; -
KW EMBL; AF348506; AAK31344.1; -
DR Coiled coil. 151 375 COILED COIL (POTENTIAL).
KW DOMAIN 387 803 COILED COIL (POTENTIAL).
FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

Query Match 4.6%; Score 268; DB 1; Length 1208;
Best Local Similarity 17.7%; Pred. No. 0.0083;
Matches 232; Conservative 219; Mismatches 445; Indels 414; Gaps 55;

Qy 50 NDRD-NRQAFEGISQLREEYSNKAIKNPTKKNQYFSDPFINKSNLDINKDNLIDVESSTKS 108
Dy 2 SERDFTQS-----PKFDENANSVISQ-----SDFLGKSLKNNDD----- 38
Qy 109 FQKFGDQRYRIFTSWSHQNDPSPKINTRSIRNFMENTIQP-----PILDDKEKAEFLK 161
Dy 39 ---YSDFR-----GSYLNKDKSSFQ-PLRN---GSYQPKGSLEFPLLOSSNK---N 80
Qy 162 SAKOSFAGIIGNOIRTDQKFMGVDFESLKERQBAENGEPTGDMWLDIFLFTFDKKQS 221
Dy 81 SDKYNGS---LGDKSPDPNSYGLSAISKQATQEALESISQ--GNDSDYDV--SKLTDLN 133
Qy 222 SDVKEAINQEPVPHVQPIATTTTD-----IOGLPPE--- 253
Dy 134 SEIDHTGELPA-----NAULTREQEKVLEKVSRENFGRIKIVCLEKRLSEMAPEQIK 188
Qy 254 --ARDLLD---ERGNFSKFTLGDMEMLDVEGV--ADIDPNYKFNQLLIHNNALSSVLMS 306
Dy 189 EAVKDNVELHAERANL-----QLQKRTESLLQKSEDKNFKLEKVDYLSKVNDVEQS 242
Qy 307 HNGIEPEKVSLLYGGNGGPGARHDWNAIVGYKQDQGNVATIIIVHMKNGSLVIAGEK 366
Dy 243 NVKVFTERIRFLENALKVQREKDSLSTEMEEDKSNEVD----- 282
Qy 367 GINNPSPYLYKEDQLTGSQALSQEEIQNKIDFMEFLAONNAKLDNLSEKE--KRT 424
Dy 283 -----YEYEROL-----QNRJD--ELSEELQVADLLTEKEDEIATUKRQ 321
Qy 425 IKDFQKDSKAYLDALGNDRIAFVSKKDTKHSALITEFGNGDLSYTLKDYG---KKADKAL 481
Dy 322 IEKENSSSAF-----ENEENSSVY-----HLQEDYALLOAKCDEFA 358
Qy 482 DREKNVTLOGSLKHGDMVFDVSNFYNTASKNPNKNGVGVTVNGV-----SHL 528
Dy 359 DRIQ-----VLTADLEKEKENQIMHESEASIGLTDMSQVHTLOQLKHANEI 406
Qy 529 EVGFNKVAIENLPDLNLAITSFVRRNLE---DKLTTKGLSPQEPANKLIDFLSSNKELV 585
Dy 407 EFLHDQISRWN--ERGNFEDIMIQFRSLEBERDVLESKLOTLEDNNSLRMTSSLGNQI 465
Qy 586 GKTUNFNKAVADAKNTGNYDEVKKAQKDLKSLKREHLEKEVEKKLESKGNKKNKAQ 645
Dy 466 ESURTQNRDEDEENHLRLKASKNSDKALAEITNIRLQEVTKELTMRKNSNDLNEIDL 525
Qy 646 AQANSQKDEITFALINKENR-----DARATAYAOQLKGIKRELSKLENVKNLKDFOK 700
Dy 526 REENEGTLKIDSITTEKDKRLINELEQRIKSYEVNVS-----ELNGTIDEYRNKLKDEET 581
Qy 701 FDEPKNG---KNKDFSKAEETLKALKGSVKDLGIN-----PEWLSKVENLNAAL 746
Dy 582 YNEVMNAFYQKNDRLRRFHESINKLQDREKLTSLNLEKKNLVLSLRETVAMLEKERE 641

Qy 747 NEFKNGKNKDFSKVTOAKSDLENSVKDVI INOKVTDKVDNLNOA-----VSVAKATG 798
Db 642 KKYLSGNKDLDN-TNLMEIINDKIS--VLQOQLTDVKDELVDSEEREEREAIVAGOKLSA 698
Qy 799 DF---SRVEQALADLKNFSEKQLAQAOKNESLNA-----RKKSEIYO-----SVK 841
Db 699 SFELMSNEKQAL-----ELKYSSLKNELINAQNLDRREELSSELSKKLFEERKIR 749
Qy 842 NGVNGTLVGNCLSOAEATTLTKNFSIDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKKKA 901
Db 750 SGSNDIDIEKN-----KEINVNSELAADLAQIRHLESDKMELDKLVHHLN--RG 796
Qy 902 GOAASLEE-----PIYAVAKKVNKIDRLNQTASGLGVVGQAA 940
Db 797 IEEANIEHAVKKRCLLLMGCDYSSVILQIVSQIEHFVNOQIQIIRSLKQEL----- 849
Qy 941 GFPLKRHD-----KVGL-----SRQELAQIDNLNQAVSEAK 979
Db 850 -----RHDFVQFSGKKEQELSRSEKFGCGTETKHDILAQRNRNVSEKMNDENAQK-- 902
Qy 980 AGFTGN-----LEOTIDKL-----KDKTK 998
Db 903 --FFSSPDRKNGLYPSEHTSKIEYLEKTIEDLKALQDELKRNLLMDDISSYNKQTTK 960
Qy 999 HNPMLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINEK----- 1043
Db 961 LOEKIKWLERERSI---LIDELESYRSNQF-----NYONNLVQDKNELEERLKEIQKELE 1012
Qy 1044 -----ATCMLTQKNPEWKLKLVNDKIVAHNVGVSPLSEYDKIGFNOKNKKDYSD 1091
Db 1013 VYNNHFMKQAEMLTSNVTDESQMLKTLREALQSKTNNIDHLSTI--LERNRKEYKSLLD 1070
Qy 1092 SF-----KFSTKLNNAVKTNSGFTQFLTNAFSTASYCYCLARENAEHGK 1136
Db 1071 DYNOLRARYKNLQSNTPOSTOSGOVESEIKGLSKLTKYLOSKCRREHSR 1120

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2002, 14:01:01 ; Search time 118.81 Seconds
(without alignments)
1412.123 Million cell updates/sec

Title: US-09-360-685A-5
Perfect score: 5867
Sequence: 1 MTNETIDQOQTEAAFNQQ.....RENAHGKKNVTKGFGQKS 1147

Scoring table: BLOSUM62
Gapop 10.0 , Gapext.0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5859	99.9	1147	2	Q9F220 helicobacte
2	5550	94.6	1179	2	Q9L5X9 helicobacte
3	5448	92.9	1179	2	Q9F222 helicobacte
4	5403	92.1	1247	2	Q9RF15 helicobacte
5	5394	91.9	1247	2	Q9F223 helicobacte
6	5352	91.2	1165	2	Q9F221 helicobacte
7	5308	90.5	1247	2	Q07910 helicobacte
8	5263	89.7	1183	2	Q86064 helicobacte
9	4741	80.8	1163	2	Q9F230 helicobacte
10	4720	80.4	1171	2	Q9F229 helicobacte
11	4703.5	80.2	1170	2	Q9F218 helicobacte
12	4700	80.1	1173	2	Q9RF16 helicobacte
13	4690.5	79.9	1176	2	Q9F231 helicobacte
14	4690.5	79.9	1176	2	Q9F228 helicobacte
15	4689	79.9	1177	2	Q9F225 helicobacte
16	4679	79.8	1173	2	Q9KHU3 helicobacte
17	4678.5	79.7	1230	2	Q9F232 helicobacte
18	4677.5	79.7	1172	2	Q9F227 helicobacte
19	4664.5	79.5	1186	2	Q9F224 helicobacte

20	4658.5	79.4	1172	2	Q9F217	09f217 helicobacte
21	4640	79.1	1227	2	Q9F219	09f219 helicobacte
22	4635.5	79.0	1172	2	Q9F226	09f226 helicobacte
23	1745.5	29.8	441	2	Q9L7K3	09l7k3 helicobacte
24	1718	29.3	408	2	Q9L7K4	09l7k4 helicobacte
25	1695.5	28.9	445	2	Q9L7K5	09l7k5 helicobacte
26	838	14.3	187	2	Q9X5H4	09x5h4 helicobacte
27	764	13.0	149	2	Q9RLY1	09rly1 helicobacte
28	763	13.0	149	2	Q9RLZ1	09rlz1 helicobacte
29	762	13.0	149	2	Q9RLX7	09rlx7 helicobacte
30	760	13.0	149	2	Q9RLZ3	09rlz3 helicobacte
31	759	12.9	149	2	Q9RLY4	09rly4 helicobacte
32	758	12.9	149	2	Q9RLZ0	09rlz0 helicobacte
33	758	12.9	149	2	Q9RLX6	09rlx6 helicobacte
34	757	12.9	149	2	Q9RLY5	09rly5 helicobacte
35	757	12.9	149	2	Q9RLX8	09rlx8 helicobacte
36	756	12.9	149	2	Q9RLZ2	09rlz2 helicobacte
37	754	12.9	149	2	Q9RLX5	09rlx5 helicobacte
38	751	12.8	149	2	Q9RLY9	09rly9 helicobacte
39	749	12.8	149	2	Q9RLY7	09rly7 helicobacte
40	749	12.8	149	2	Q9RLY6	09rly6 helicobacte
41	749	12.8	149	2	Q9RLY3	09rly3 helicobacte
42	748	12.7	149	2	Q9RLY2	09rly2 helicobacte
43	746	12.7	149	2	Q9RLY8	09rly8 helicobacte
44	731	12.5	149	2	Q9RLX9	09rlx9 helicobacte
45	718	12.2	149	2	Q9RLY0	09rly0 helicobacte

ALIGNMENTS

RESULT 1

ID Q9F220 PRELIMINARY; PRT: 1147 AA.
AC Q9F220;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE CACA.
GN CACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC11637;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H., Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from Japanese and non-Japanese isolates."
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015416; BAB20926.1;
SQ SEQUENCE 1147 AA; 128003 MW; D740A7275A9B141D CRC64;

Query Match 99.9%; Score 5859; DB 2; Length 1147;
Best Local Similarity 99.9%; Pred. No. 9,9e-238;
Matches 1146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNETIDQOQTEAAFNQQFINNQVAFKLVNDNAVAYDDPQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQOQTEAAFNQQFINNQVAFKLVNDNAVAYDDPQKPIVDKNDNRDNRQAFEG 60
QY 61 ISOLREESNKAIKNPTKKNQYFSDFINKSNLIDVESSTKSKFKGQDQRYRIF 120
Db 61 ISOLREESNKAIKNPTKKNQYFSDFINKSNLIDVESSTKSKFKGQDQRYRIF 120
QY 121 TSWVSHQNDPSKINRSIRNFNMENIIQPIILDDKKAELFKSAKOSFAGIIIGNIQTQD 180
Db 121 TSWVSHQNDPSKINRSIRNFNMENIIQPIILDDKKAELFKSAKOSFAGIIIGNIQTQD 180
QY 181 KPMGVFDSLSKQERQAEKNGEPTGGDWLDIFLSFIDFKKQSSDVKEAINQEPVPHVQDDI 240

Db 181 KFMGVFDESLEKQAEKNGEPTGGDWLDFLSFTFDDKSSDVKEATNQEPVPHVQDI 240
Qy 241 ATTTTDTIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHHNNALS 300
Db 241 ATTTTDTIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHHNNALS 300
Qy 301 SVLMGSHNGIEPEKVSLLYGNGGPGGARHDMNATVGYKDDQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGGARHDMNATVGYKDDQGNVATTINVHMKNKSGSLV 360
Qy 361 IAGGEGKINNPFYLYKEDQLTGSGORALSQBEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPFYLYKEDQLTGSGORALSQBEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Qy 421 FRTEIKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYCKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYCKKADKA 480
Qy 481 LDREKNVTLOGSLKHGDMVDFVSNFYKYNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVDFVSNFYKYNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Qy 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTGLNFNKAADAKN 600
Db 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTGLNFNKAADAKN 600
Qy 601 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Qy 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTOAKSDLENSVKDVIINQKV 780
Qy 781 TDKVDNLNQAVSAKATGDFSRVFOALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSAKATGDFSRVFOALADLNKFSKEQLAQQAQAKNESLNARKKSEIYQSV 840
Qy 841 KNGVGTGLVGNLSQAEATTLSKNFSDIKKELNKLNFNNNNNGLKNEPIYAKVKKK 900
Db 841 KNGVGTGLVGNLSQAEATTLSKNFSDIKKELNKLNFNNNNNGLKNEPIYAKVKKK 900
Qy 901 AGQAASLEPIYAQVAKVKNNAKIDRLNQIASGLGVGQAGFPPLKRHDVDDLSKVGLSR 960
Db 901 AGQAASLEPIYAQVAKVKNNAKIDRLNQIASGLGVGQAGFPPLKRHDVDDLSKVGLSR 960
Qy 961 NOELAQKIDNLNQAVSEAKAGFGNLEQITDKLKDSTKHNNPMLWVESAKKVPASLSAKL 1020
Db 961 NOELAQKIDNLNQAVSEAKAGFGNLEQITDKLKDSTKHNNPMLWVESAKKVPASLSAKL 1020
Qy 1021 DNYATNSHIRINSINKNGAINKEKATGMLTKQNPWLKLVNDKIVAHNVGSPVLEIDYKIG 1080
Db 1021 DNYATNSHIRINSINKNGAINKEKATGMLTKQNPWLKLVNDKIVAHNVGSPVLEIDYKIG 1080
Qy 1081 FNQNMKDYSDSFSTKLNNAVKDNTSGFTQFLTNAPSTASYCYCLARENAEHGKKNVT 1140
Db 1081 FNQNMKDYSDSFSTKLNNAVKDNTSGFTQFLTNAPSTASYCYCLARENAEHGKKNVT 1140
Qy 1141 KGGFQKS 1147
Db 1141 KGGFQKS 1147

RESULT 2
Q9L5X9 PRELIMINARY; PRT: 1179 AA.
ID Q9L5X9;
AC Q9L5X9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CYTOTOXIN-ASSOCIATED PROTEIN CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN (1) SEQUENCE FROM N.A.
RP STRAIN=SS1;
RC Zhang J.2.;
RA "cagA gene of Helicobacter pylori (SS1 strain).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN (2) SEQUENCE FROM N.A.
RP STRAIN=SS1;
RC Zhou J.C.; Zhang J.2.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF247651; AAF63759.1;
DR EMBL: 1179 AA; 131297 MW; 7A9112F58B749787 CRC64;
SQ SEQUENCE 1179 AA; 131297 MW; 7A9112F58B749787 CRC64;

Query Match 94.6%; Score 5550; DB 2; Length 1179;
Best Local Similarity 92.3%; Pred. No. 8.8e-225;
Matches 1090; Conservative 19; Mismatches 36; Indels 36; Gaps 3;

Qy 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNNAVASYDPDQKPIVDKNDNRDRAPEG 60
Db 1 MTNETINQOPQTEAAFNPOQFINNLOVAFKVDNNAVASYDPDQKPIVDKNDNRDRAPEG 60

Qy 61 ISQLEESNKAINKPTKKNOVFSDFINKSNDLKNLIDVESSTKSFQKFGDQRYIF 120
Db 61 ISQLEESNKAINKPTKKNOVFSDFINKSNDLKNLIDVESSTKSFQKFGDQRYIF 120

Qy 121 TSWVSHQNDPSKINTSRINFEMENIQQPILDDEKAEFLKSQKSAFAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTSRINFEMENIQQPILDDEKAEFLKSQKSAFAGIIGNQIRTDQ 180

Qy 181 KFMGVFDESLEKQAEKNGEPTGGDWLDFLSFTFDDKSSDVKEATNQEPVPHVQDI 240
Db 181 KFMGVFDESLEKQAEKNGEPTGGDWLDFLSFTFDDKSSDVKEATNQEPVPHVQDI 240

Qy 241 ATTTTDTIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHHNNALS 300
Db 241 ATTTTDTIOGLPEARDLDDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLHHNNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYGNGGPGGARHDMNATVGYKDDQGNVATTINVHMKNKSGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYGNGGPGGARHDMNATVGYKDDQGNVATTINVHMKNKSGSLV 360

Qy 361 IAGGEGKINNPFYLYKEDQLTGSGORALSQBEIQNKIDFMEFLAQNNAKLDNLSEKEK 420
Db 361 IAGGEGKINNPFYLYKEDQLTGSGORALSQBEIQNKIDFMEFLAQNNAKLDNLSEKEK 420

Qy 421 FRTEIKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYCKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRITAFVSKDTKHSALITEFGNGDLSYTLKDYCKKADKA 480

Qy 481 LDREKNVTLOGSLKHGDMVDFVSNFYKYNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540
Db 481 LDREKNVTLOGSLKHGDMVDFVSNFYKYNASKPNKNGVGTNGVSHLEVGFKVAIFNL 540

Qy 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTGLNFNKAADAKN 600
Db 541 PDLNNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTGLNFNKAADAKN 600

Qy 601 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660
Db 601 TGNVDEVKKAOKDLEKSLRKRHEHLEKEVEKKLESKSGNKNKMEAKAQAANSQKDEIFALIN 660

Qy 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720
Db 661 KEANRDARAIYAQNLIKRELSKLENVKNLKDPSKDFEFKNGKNKDFSKAEETLK 720

```
Db 661 KEANRARAIAATNLRGVRKRELSKLENVKNLKDPSFDEPKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINOKV 780
Db 721 TLKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVIINOKI 780
QY 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQLAQOAKNESINARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQLAQOAKNEDFTGKNKSAIYQSV 840
QY 841 KNGVNGTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNGTLVGNGLSKAEATTLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
QY 901 AGQAASLEPIYAAQVAKVNAKIDRLNQIASGLGVQQAAGFPLKRHDKVD-----951
Db 901 AGQVSPSEPIYAAQVAKVNAKIDQLNQAAASGLGVQ-AGFPLKRHDKVDLDSKVGRSV 959
QY 952 -----DLSKVGLSRNQELAQKIDNLNQAVSEAKAGFFGNL 986
Db 960 SPEPIYATIDDLGGPFLPKHAKVEDLSKVGLSREQELTOKIDNLNQAVSEAKAGFFGNL 1019
QY 987 EOTIDKLDSTKHNPMNLWVESAKKVPASLSAKLDNYATNSHIRINSNIKNGAINEKATG 1046
Db 1020 EQTIDNLKSTKHNPMNLWESAKKVPASLSAKLDNYATNSHTRINSNIQNGAINEKATG 1079
QY 1047 MLTQKNPEWLKLVNDKIVAHNVGSPVSEYDKIGFNOKNMKDYSDSFSTKLNNAVKT 1106
Db 1080 MLTQKNPEWLKLVNDKIVAHNVGSPVSEYDKIGFNOKNMKDYSDSFSTKLNNAVKT 1139
QY 1107 NSGFTQFLTNAFSTASYCYCLARENAEHGKKNVTKGGFOKS 1147
Db 1140 RSGETQFLANAFST-GYYCLAEENAEHGIENVNTKGGFOKS 1179

RESULT 3
Q9F222 ID Q9F222 PRELIMINARY; PRT; 1179 AA.
AC Q9F222;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43579;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015414; BAB20924.1; -
SQ SEQUENCE 1179 AA; 131686 MW; 7D0ADC34E644634 CRC64;

Query Match 92.9%; Score 5448; DB 2; Length 1179;
Best Local Similarity 90.9%; Pred. No. 1.6e-220;
Matches 1073; Conservative 23; Mismatches 50; Indels 34; Gaps 2;

QY 1 MTNETIQOQFOTEAAFPQOFINNLQVAFKLVNDNAVASYDPDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIQOQFOTEAAFPQOFINNLQVAFKLVNDNAVASYDPDQKPIVDKNDNRDNRHAFEG 60
QY 61 ISOLREESYNAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFKQFGDQRYRIF 120
Db 61 ISOLREESYNAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFKQFGDQRYRIF 120
QY 121 TSWVSHQNDPSKINTRSRNFMENIQQPIPDDEKAEFLASAKQSFAGIIVGNQIRTDQ 180
```

```
Q9RF15
ID Q9RF15 PRELIMINARY; PRT; 1247 AA.
AC Q9RF15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11637;
RA Ito Y., Azuma T.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF202973; AAF17598.1; -.
SQ SEQUENCE 1247 AA; 139042 MW; 28DEC5888F9D041B CRC64;

Query Match 92.1%; Score 5403; DB 2; Length 1247;
Best Local Similarity 85.5%; Pred. No. 1.3e-218;
Matches 1067; Conservative 35; Mismatches 44; Indels 102; Gaps 2;

Qy 1 MTNETIDQPOQTEAAAFNPQQFNNLQVAFKVDNAVASYPDPQPIVDKNDNRNRQAFEG 60
Db 1 MTNETIDQPOQTEAAAFNPQQFNNLQVAFKVDNAVASYPDPQPIVDKNDNRNRQAFEG 60

Qy 61 ISQREESYNAIKNPYKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQFGQORYRIF 120
Db 61 ISQREESYNAIKNPYKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQFGQORYRIF 120

Qy 121 TSWYSHQNDPSKINTSRIRNMENTIQQPIIDDKKEAEFLKSAKSPAGIIIGNOIRTDQ 180
Db 121 TSWYSHQNDPSKINTSRIRNMENTIQQPIIDDKKEAEFLKSAKSPAGIIIGNOIRTDQ 180

Qy 181 KFMGVFDESLKERQEAENKGEPTGGDWLDIFLSIFDKQSSDYKKEAINQEPVPHVQPD 240
Db 181 KFMGVFDEFLKERQEAENKGEPTGGDWLDIFLSEFVENKEQSSDYKKEAINQEPVPHVQPD 240

Qy 241 ATTTTIDQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIPNFKFNOLLHNALS 300
Db 241 ATTTTIDQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIPNFKFNOLLHNALS 300

Qy 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAVTVGYKQOQDNVATLINVHMKGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAVTVGYKQOQDNVATLINVHMKGSLV 360

Qy 361 IAGGEGINNPSFYLYKEDQITQSORALSQBEIQNKIDFMFLAQNNAKLDNLSEKEK 420
Db 361 IAGGEGINNPSFYLYKEDQITQSORALSQBEIRNKIDFMFLAQNNAKLDNLSEKEK 420

Qy 421 FRTIKDFQKDSKAYLDALGNDRITAFYSKDKTKHSALITEFGNGDLSYTLKDYKKAADKA 480
Db 421 FQNEIEDFQKDSKAYLDALGNDRITAFYSKDKTKHSALITEFGNGDLSYTLKDYKKAADKA 480

Qy 481 LDREKNVTLOGSLKHGDMFVDYSNEFYTNASKNPNGVGTNGVSHLEVGFNKAVALFNL 540
Db 481 LDREKNVTLOGSLKHGDMFVDYSNEFYTNASKNPNGVGTNGVSHLEVGFNKAVALFNL 540

Qy 541 PDLNNAITSVRRNLEDKLTKTKSQEANKLIKDFLSSNKNELVGKTLNFKNAVADAKN 600
Db 541 PDLNNAITSVRRNLEDKLTKTKSQEANKLIKDFLSSNKNELVGKTLNFKNAVADAKN 600

Qy 601 TGNVDEYKKAQKDLKSLRREHLEKEVEKKLESKSNKNKMEAKAANSOKDIFALIN 660
Db 601 TGNVDEYKKAQKDLKSLRREHLEKEVEKKLESKSNKNKMEAKAANSOKDIFALIN 660

Qy 661 KEANRDRATAYAGNLKIKRELSDKLENNVKNLKDQKSFDEPKNGKNKDFSKAETLK 720
Db 661 KEANRDRATAYAGNLKIKRELSDKLENNVKNLKDQKSFDEPKNGKNKDFSKAETLK 720
```

```
Qy 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSKVDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSKVDVIINQKI 780

Qy 781 TDKVDNLNQAVSVAKATGDFSRVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSMAKATGDFSRVEQALADLKNFSKEQLAQQAQTKNEFNVGKSEIYQSV 840

Qy 841 KNGVNGTILVGNLSQAEATLTSKNFSDIKKELNAKLNFNNNNNNNKNEPIYAKVNNKK 900
Db 841 KNGVNGTILVGNLSGIEATLAKNFSDIKKELNEKFKNPNNNNNNNNGLENEPIYAKVNNKK 900

Qy 901 AQQAASLEPIYAQVAKVNAKIDRLNQIASGLGVVQGA----- 939
Db 901 TQOVASPEPIYAQVAKVNAKIDRLNQAAASGLGVGQAGFPLKRHDKVDDLSKYGRSVS 960

Qy 940 ----- 939
Db 940 ----- 939

Qy 961 PEPIYATIDDLGGPPPLKRHDKVDDLSKYGRSVSPEPIYATIDDLGGPPPLKRHDKVDDL 1020
Db 961 -----AGFPLKRHDKVDDLSKVGLSRNOELAQKIDNLNOAVSEAK 979

Qy 1021 SKVGRSVSPEPIYATIDDLGGPPPLKRHDKVDDLSKVGLSRNOELAQKIDNLNOAVSEAK 1080
Db 1021 SKVGRSVSPEPIYATIDDLGGPPPLKRHDKVDDLSKVGLSRNOELAQKIDNLNOAVSEAK 1080

Qy 980 AGFEGLNLEQITDKLKDSTKHNPMLNWSAKKVPASLASAKLDNYATNSHIRINSNIKGA 1039
Db 1081 AGFESNLEQITDKLKDSTKYNSVNLWVESAKKVPASLASAKLDNYATNSHTRINSNIQGA 1140

Qy 1040 INEKATGMLTOKNPWLKLVNDKIVAHNVGSLSEYDKIGNQKNMKDYSDSFSTKL 1099
Db 1141 INEKATGMLTOKNPWLKLVNDKIVAHNVGSLSEYDKIGNQKNMKDYSDSFSTKL 1200

Qy 1100 NNAVDTNSGTQPLTNAFSTASYCLARENAEHGKIKVNTKGGFQKS 1147
Db 1201 NNAVDTNSGTQPLTNAFST-GYISLARENAEHGKIKVNTKGGFQKS 1247

RESULT 5
Q9F223 PRELIMINARY; PRT; 1247 AA.
ID Q9F223
AC Q9F223;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43526;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J Gastroenterol. 35:890-897(2000).
DR EMBL; AB015413; BAB20923.1; -.
SQ SEQUENCE 1247 AA; 139064 MW; 121E68825E5C9228 CRC64;
```

```
Query Match 91.9%; Score 5394; DB 2; Length 1247;
Best Local Similarity 85.3%; Pred. No. 3.2e-218;
Matches 1065; Conservative 35; Mismatches 46; Indels 102; Gaps 2;
```

```
Qy 1 MTNETIDQPOQTEAAAFNPQQFNNLQVAFKVDNAVASYPDPQPIVDKNDNRNRQAFEG 60
Db 1 MTNETIDQPOQTEAAAFNPQQFNNLQVAFKVDNAVASYPDPQPIVDKNDNRNRQAFEG 60

Qy 61 ISQREESYNAIKNPYKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQFGQORYRIF 120
Db 61 ISQREESYNAIKNPYKKNQYFSDFINKSNLDLNKNDLIDVESSTKSFQFGQORYRIF 120
```

```
QY 121 TSWVSHQNDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
QY 181 KFMGVDESLEKROEAEKNGEPTGGDWLDIFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVDEFLEKROEAEKNGEPTGGDWLDIFLSFVFNKEQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDQGLPPARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTHIQLGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKQDQGNVATIIINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNATVGYKQDQGNVATIIINVHMKNGSLV 360
QY 361 IAGEGKGINNPSFYLKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGEGKGINNPSFYLKEDQLTGSORALSOEIRNKIDPMEFLAONNAKLDNLSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFNGDLSYTLKDYGKKADKA 480
Db 421 FONEIEDFOKDSKAYLDALGNDRIFAUSKDKTKHSALITEFNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFDVSNFKYTNASKNPKNKGVTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGNKLKHDGVMFVNTSNFKYTNASKNPKNKGVTNGVSHLDAGFSKVAIFNL 540
QY 541 PDNLNLAITSFVRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDNLNLAITSFVRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
QY 601 TGNVDEKKAQKDLERSLKRREHLEKEVEKKLESKSGKNKMEAKAQAQNSOKDEIFALIN 660
Db 601 TGNVDEKKAQKDLERSLKRREHLEKEVEKKLESKSGKNKMEAKAQAQNSOKDEIFALIN 660
QY 661 KEANRARAIAQNLIKIGIKRELSDKLENVKNLKDPSDFEFKNGKNDKFSKAEETLK 720
Db 661 KEANRARAIAQNLIKIGIKRELSDKLENVKNLKDPSDFEFKNGKNDKFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKOPSKVTOAKSDLENSVKDVIINQKV 780
Db 721 ALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKOPSKVTOAKSDLENSVKDVIINQKV 780
QY 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQALAAQAKNESLNARKKSEIYQSV 840
Db 781 TDKVDNLNQAVSAKATGDFSRVEQALADLNKFSKEQALAAQAKNESLNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSQABATLSKNFSDIKKELNAKLGNFNNNNNGLKNEPIYAKVNKK 900
Db 841 KNGVNGTLVGNGLSGTEATALAKNFSDIKKELNEKFNFNNNNNGLNEPIYAKVNKK 900
QY 901 AGQAASLEPIYAAQVAKVNAKIDRLNQIASGLGVQCA----- 939
Db 901 TGQVASPEEPIYAAQVAKVNAKIDRLNQIASGLGVQAGPFLKRHDKVDLDSKVGSRVS 960
QY 940 ----- 939
Db 940 ----- 939
QY 961 PEPIYATIDDLGGPFLKRHDKVDDLSKVGSRVSPEPIYATIDDLGGPFLKRHDKVDDL 1020
Db 961 PEPIYATIDDLGGPFLKRHDKVDDLSKVGSRVSPEPIYATIDDLGGPFLKRHDKVDDL 1020
QY 940 -----AGPFLKRHDKVDDLKSVGLSRNQELAQKIDNLNQAVSEAK 979
Db 1021 SKVGRSVSPEPIYATIDDLGGPFLKRHDKVDDLKSVGLSRNQELAQKIDNLNQAVSEAK 1080
QY 980 AGFEFNLEQIDKLDKOSTKNPNLWVESAKVYPASLSAKLDNYATNSHTRINSINQGA 1039
Db 1081 AGFEFNLEQIDKLDKOSTKYNSLWVESAKVYPASLSAKLDNYATNSHTRINSINQGA 1140
QY 1040 INEKATGMLTKQNPWELKLVNDKIVAHNVGSVPLSEYDKTGFQKNMKDYSDSFKFSTKL 1099
Db 1141 INEKATGMLTKQNPWELKLVNDKIVAHNVGSVPLSEYDKTGFQKNMKDYSDSFKFSTKL 1200
```

```
QY 1100 NNAVKDTNSGFTQPLTNAPSTASYCYCLARENAEHGKIKNVNTKGGFOKS 1147
Db 1201 NNAVKDVKSSTOFLANAFST-GYISLARENAEHGKIKNVNTKGGFOKS 1247

RESULT 6
Q9F221 PRELIMINARY; PRT: 1165 AA.
AC Q9F221;
DT 01-MAR-2001 (Tremblurel. 16, Created)
DT 01-MAR-2001 (Tremblurel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblurel. 16, Last annotation update)
DE .CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC49503;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RT J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015415; BAB20925.1; -.
SQ SEQUENCE 1165 AA; 130140 MW; 99C316921E649C02 CRC64;

Query Match 91.2%; Score 5352; DB 2; Length 1165;
Best Local Similarity 90.4%; Pred. No. 1.7e-216;
Matches 1057; Conservative 28; Mismatches 58; Indels 26; Gaps 7;

QY 1 MTNETIDQQPOTEAAPNQPFINNLOVAFLKVDNAVASYDPDQKPIVDKNDNRDNRQAFEG 60
Db 1 MTNETIDQQPOTEAAPNQPFINNLOVAFLKVDNAVASYDPDQKPIVDKNDNRDNRQAFEG 60
QY 61 ISQLEEYSNKAINKPTKKNQYFSDFINKSNDLINKDLIDVESSTKSFKQFGDQRYRIF 120
Db 61 ISQLEEYSNKAINKPTKKNQYFSDFINKSNDLINKDLIDVESSTKSFKQFGDQRYRIF 120
QY 121 TSWVSHQNDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
Db 121 TSWVSHQNDPSKINTRSIRNFEMENIIQPPILDDKEAEFLKSAKQSPAGIIGNQIRTDQ 180
QY 181 KFMGVDESLEKROEAEKNGEPTGGDWLDIFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
Db 181 KFMGVDESLEKROEAEKNGEPTGGDWLDIFLSFIPDKKQSSDVKEAINQEPVPHVQPD 240
QY 241 ATTTTIDQGLPPARDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
Db 241 ATTTTHIQLGLPPESRDLLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYGGNGPGARHDWNATVGYKQDQGNVATIIINVHMKNGSLV 360
Db 301 SVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNATVGYKQDQGNVATIIINVHMKNGSLV 360
QY 361 IAGEGKGINNPSFYLKEDQLTGSORALSOEIQNKIDPMEFLAONNAKLDNLSEKKEK 420
Db 361 IAGEGKGINNPSFYLKEDQLTGSORALSOEIRNKIDPMEFLAONNAKLDNLSEKKEK 420
QY 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFNGDLSYTLKDYGKKADKA 480
Db 421 FRTEIKDFQKDSKAYLDALGNDRIFAUSKDKTKHSALITEFNGDLSYTLKDYGKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMVFDVSNFKYTNASKNPKNKGVTNGVSHLEVGFNKVAIFNL 540
Db 481 LDREKNVTLOGNKLKHDGVMFVNTSNFKYTNASKNPKNKGVTNGVSHLDAGFSKVAIFNF 540
QY 541 PDNLNLAITSFVRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
Db 541 PDNLNLAITSFVRNLEDKLTTKGLSPOEANKLIKDFLSSNKELVGKTLNFKNAVADAKN 600
```

```

QY 601 TCNYDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
DB 601 TCNYDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
QY 661 KEANDARAIAAYAOHLKIGIKRELSDKLENVKNLKDQKDFDEPKNGKNKDFSKAEETLK 720
DB 661 KEANREARAIYAYAOHLKIGIKRELHDNLNINKDLNFKSDFEKNKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTQAQSDLENSKVDIINQKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTQAQSDLENSKVDIINQKV 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
DB 781 TDKVDNLNQAVSVAKATGDFSGVEQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSOAEATTLKSNFSDIKKELNAKLNKFNNNNNGLKN--EPIYAKVKN 898
DB 841 KNSVNTKLVGNGLSGIEATLAKNFSQIKKELNEKFNKFN--NNNNNGFNKFNSTEPIYAKVKN 899
QY 899 KKQAQASLEPIYAAQAKV-----NAKIDRLNQAISGL-----GVVGOA 939
DB 900 KKTGOVASPREPIYAAQVKNKKKAGAAAGPPLKRHDKVDLDSKV--GLSASPEPIYATIDDL 957
QY 940 AG-FPLKRHDKVDLKSGLSRNQLAQKIDNLAQVSEAKAGFFGNLEQITDKLKDSTK 998
DB 958 GGPEPLKRHDKVDLKSGLSRNQLAQKIDNLAQVSEAKAGFFGNLEQITDKLKDFTK 1017
QY 999 HNPMLWVESAKVPASLSAKLDNYATNSHTRINSNTKNGAINKATCMLTQKNPEWLK 1058
DB 1018 HNPMLWVESAKVPASLSAKLDNYATNSHTRINSNTKNGAINKATCMLTQKNPEWLK 1077
QY 1059 VNDKIVAHNVGSVPLSEYDKTIGFQKNMKDYSDSKFESTKLNNAVKDTNSGFTQFLTNAF 1118
DB 1078 VNDKIVAHNVGSVPLSEYDKTIGFQKNMKDYSDSKFESTKLNNAVKDTNSGFTQFLTNAF 1137
QY 1119 STASYICLARENAEHGIKNNVNTKGGFQKS 1147
DB 1138 ST-GYYCLARENAEHGIKNNVNTKGGFQKS 1165

RESULT 7
ID 007910 PRELIMINARY; PRT; 1247 AA.
AC 007910;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CAGA (CYTOTOXIN ASSOCIATED PROTEIN A).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43526;
RA Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RA Shiratori Y., Omata M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43526;
RA Maeda S., Kanai F., Ogura K., Yoshida H., Ikenoue T., Kawabe T.,
RA Shiratori Y., Omata M.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 335-415 FROM N.A.
RC STRAIN=NCTC11637;
RA Auerman M., Azuma T., Berg D.E., Ito Y., Morelli G., Pan Z.J.,
RA Suerbaum S., Thompson S., van der Ende A., van Doorn L.J.;
RT "Recombination and clonal groupings within Helicobacter pylori from

```

```

RT different geographic regions.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001357; AAB58747.1; -
DR EMBL; AB003397; BAA19853.1; -
DR EMBL; AJ239729; CAB37821.1; -
SQ SEQUENCE 1247 AA; 138819 MW; 172C76AE9F9BD29E CRC64;

Query Match          90.5%; Score 5308; DB 2; Length 1247;
Best Local Similarity 84.1%; Pred. No. 1.3e-214;
Matches 1051; Conservative 41; Mismatches 53; Indels 104; Gaps 3;

QY 1 MTNETIDQOPQTEAAFNPOQFINNLOVAFKVDNNAVSYDPQKPIVDKNDNRNQAPEG 60
DB 1 MTNETIAQOPQTEAAFNPOQFINNLOVAFKVDNNAVSYDPQKPIVDKNDNRNQAPEG 60
QY 61 ISOLREESNKAIRNPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDORYRIF 120
DB 61 ISOLREESNKAIRNPTTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFQFGDORYRIF 120
QY 121 TSWYSHONDPSKINTSRIRNFEMENIOPPIIDDKKEAEFLKSAKOSFAGIIGNOIRTDQ 180
DB 121 TSWYSHONDPSKINTSRIRNFEMENIOPPIIDDKKEAEFLKSAKOSFAGIIGNOIRTDQ 180
QY 181 KFMGVFDESLEKQERAEKNGEPTGGDWLIDFLSFIFDKKQSSDVKEAINQBPVPHVQPD 240
DB 181 KFMGVFDESLEKQERAEKNGEPTGGDWLIDFLSFIFDKKQSSDVKEAINQBPVPHVQPD 240
QY 241 ATTTTDTQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIDPNYKFNOLLINNNALS 300
DB 241 ATTTTDTQGLPEARDLLDERGNSKFTLGDMEMLDVEGVADIDPNYKFNOLLINNNALS 300
QY 301 SVLMGSHNGIEPEKVSLLYCGNGGPGARHDWNATVGYKQDQGNVATINVMKNGSLV 360
DB 301 SVLMGSHNGIEPEKVSLLYCGNGGPGARHDWNATVGYKQDQGNVATINVMKNGSLV 360
QY 361 IAGKEGINNPSFYLYKEDQITQSORALSQBEIONKIDFMEFLAQNNKALNLSKEKEK 420
DB 361 IAGKEGINNPSFYLYKEDQITQSORALSQBEIONKIDFMEFLAQNNKALNLSKEKEK 420
QY 421 FRTEIKOFQKSKAYLDALGNDRITAFVSKKDTKISALITFEFGNGDLSTLYDKYKKADKA 480
DB 421 FQNEIEDFQKSKAYLDALGNDRITAFVSKKDTKISALITFEFGNGDLSTLYDKYKKADKA 480
QY 481 LDREKNVTLOGSLKHGDMFVYDYNFYKYNATKNPKNGVGTNGVSHLEVGNKVAIFNL 540
DB 481 LDREKNVTLOGSLKHGDMFVYDYNFYKYNATKNPKNGVGTNGVSHLEVGNKVAIFNL 540
QY 541 PDLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTTLNFKAVADAKN 600
DB 541 PDLNLAITSFVRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVGTTLNFKAVADAKN 600
QY 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
DB 601 TGNVDEVKKAQKDLKSLRKHLEKEVEKKLESKSGNKNKMEAKQAQANSOKDEIFALIN 660
QY 661 KEANDARAIAAYAOHLKIGIKRELSDKLENVKNLKDQKDFDEPKNGKNKDFSKAEETLK 720
DB 661 KEANREARAIYAYAOHLKIGIKRELHDNLNINKDLNFKSDFEKNKNGKNKDFSKAEETLK 720
QY 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTQAQSDLENSKVDIINQKV 780
DB 721 ALKGSVKDLGINPEWISKVENLNALNEFKNGKNKDFSKVTQAQSDLENSKVDIINQKV 780
QY 781 TDKVDNLNQAVSVAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
DB 781 TDKVDNLNQAVSVAKATGDFSRVQALADLKNFSKEQLAQQAQKNESLNARKKSEIYQSV 840
QY 841 KNGVNGTLVGNGLSOAEATTLKSNFSDIKKELNAKLNKFNNNNNGLKN--EPIYAKVKN 900
DB 841 KNSVNTKLVGNGLSGIEATLAKNFSQIKKELNEKFNKFN--NNNNNGFNKFNSTEPIYAKVKN 900
QY 901 AGQAASLEPIYAAQVAKVNAKIDRLNQAISGLGVVGOAAGFPPL----- 944

```

```
Db 901 TGVASPEEPIYAQVAKVNAKIDRLNQAASGLGVGQ-AGFPLKRHDKVDLDSKVGSRV 959
Qy 945 -----KRDHKVDLDSKVGSRVSPPIYATIDDLGGPPFLKRHDKVD 1019
Db 960 SPEPIYATIDDLGGPPSKRHDKVDLDSKVGSRVSPPIYATIDDLGGPPFLKRHDKVD 1019
Qy 945 -----KRDHKVDLDSKVGSRVSPPIYATIDDLGGPPFLKRHDKVD 1019
Db 1020 LSKVGSRVSPPIYATIDDLGGPPSKRHDKVDLDSKVGSRVSPPIYATIDDLGGPPFLKRHDKVD 1019
Qy 979 KAGFFGLEGOTIDKLKSTKHNPMLNWSAKKVPASLSAKLDNATNSHTRNSLNKG 1038
Db 1080 KAGFFSLEGOTIDKLKSTKHNPMLNWSAKKVPASLSAKLDNATNSHTRNSLNKG 1139
Qy 1039 AINEKATGMLTOKNPEWLKLVNDKIYAHNVGSPVLSYDKIGFNQNMKDYSDSFKESTK 1098
Db 1140 AINEKATGMLTOKNPEWLKLVNDKIYAHNVGSPVLSYDKIGFNQNMKDYSDSFKESTK 1199
Qy 1099 LNNAVKDTNSGFTQFLNAFSTASYICLARENAEHGKKNVTNKGFGOKS 1147
Db 1200 LNNAVKDTNSGFTQFLNAFST-GYYSLARENAEHGKKNVTNKGFGOKS 1247

RESULT 8
ID 086064 PRELIMINARY; PRT: 1183 AA.
AC 086064;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CYTOXIN ASSOCIATED PROTEIN III.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=15818;
RA Graessle S., Redl B., Gapp G.;
RT "Molecular characterization of the cytotoxin associated gene III
(cagAIII) of Helicobacter pylori.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083352; AAC32880.1;
SQ SEQUENCE 1183 AA; 132625 MW; 6396BD6A651A4AB7 CRC64;

Query Match 89.7%; Score 5263; DB 2; Length 1183;
Best Local Similarity 87.3%; Pred. No. 9.1e-213;
Matches 1034; Conservative 38; Mismatches 74; Indels 38; Gaps 3;

Qy 1 MTNETIDQOQTEAARNPQOQFINNLOVAFKLVNDNAVASYDDQKPIVDKNDNRQAFEG 60
Db 1 MANETINQOQTEAARNPQOQFINNLOVAFKLVNDNAVASYDDQKPIVDKNDNRQAFEG 60
Qy 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQRYIF 120
Db 61 ISQLEEYSNKAIKNPTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQRYIF 120
Qy 121 TSWVSHONDPSKINTRIRNFMENIOPPIDDKKEAEFLKSAKQSPAGIIGNQRITDQ 180
Db 121 TSWVSHONDPSKINTRIRNFMENIOPPIDDKKEAEFLKSAKQSPAGIIGNQRITDQ 180
Qy 181 KFMGVFDESLEKROEAKNCEPTGGDWLDIFLSPFDKQSSDVKEAINEQPPVPHVQPD 240
Db 181 KFMGVFDESLEKROEAKNCEPTGGDWLDIFLSPFDKQSSDVKEAINEQPPVPHVQPD 240
Qy 241 ATTTTIDQGLPPEARDLIDRGNFSKFTLGDMEMLDVEGVADIDPNYKFNQ-LLIHNNAL 299
Db 241 ATTTTIDQGLPPEARDLIDRGNFSKFTLGDMEMLDVEGVADIDPNYKFNQ-LLIHNNAL 300
Qy 300 SSVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAATVGYKDQGGNNVATIIINVHMKNGSGL 359
```

```
Db 301 SSVLMGSHNGIEPEKVSLLYAGNGGFGAKHDWNAATVGYKDQGGNNVATIIINVHMKNGSGL 360
Qy 360 VIAGGERGINNPPSYLYKEDQLTGSRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEKE 419
Db 361 VIAGGERGINNPPSYLYKEDQLTGSRALSQEEIQNKIDFMEFLAQNNAKLDNLSEKEKE 420
Qy 420 KFRTEIKDFQKDSKAYLDALGNDRIDAFVSKKDTKHSALITFEGNGDLSYTLKDYGGKADK 479
Db 421 KFRTEIKDFQKDSKAYLDALGNDRIDAFVSKKDTKHSALITFEGNGDLSYTLKDYGGKADK 480
Qy 480 ALDREKNTVLOGSLKHGDMVVDYSNEKYTNASKPNKNGVGTNCGVSHLEGVFNKVAIFN 539
Db 481 ALDREKNTVLOGSLKHGDMVVDYSNEKYTNASKPNKNGVGTNCGVSHLEGVFNKVAIFN 540
Qy 540 LPDNLNLAITSFVRRNLEDKLTTRKGLSPQEAANKLIKDFLSSNKELVGLTKLNFNAKAVADK 599
Db 541 LPDNLNLAITSFVRRNLEDKLTTRKGLSPQEAANKLIKDFLSSNKELVGLTKLNFNAKAVADK 600
Qy 600 NTGNYDEVKKAQKDLKSLRREHLEKREVEKKSKSNKNKMEAKAQAANSOKDEIFALI 659
Db 601 NTGNYDEVKKAQKDLKSLRREHLEKREVEKKSKSNKNKMEAKAQAANSOKDEIFALI 660
Qy 660 NKEANRDARATAYAQNLLGKIKRELSDKLENVKNLKDQKSFDEPKNGKNDKFSKABETL 719
Db 661 NKEASKEARAAAYVONLKGIRMEISDKLENINKDLKDFKSPDFPKNGKNDKFSKABETL 720
Qy 720 KALGSKVDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDIINOK 779
Db 721 KALGSKVDLGINPEWISKVENLAALNEFKNGKNKDFSKVTOAKSDLENSKVDIINOK 780
Qy 780 VTDKVDNLNQAASVAKATGDFSRVEQALADLKNFSKEOLAQAQKNEKSLNARKKSEIYOS 839
Db 781 ITDKVDNLNQAASVAKATGDFSRVEQALADLKNFSKEOLAQAQKNEKSLNARKKSEIYOS 840
Qy 840 VKNVNGTLVNGSLGSGEATLAKKFSKIKELNEKFNKNNNNNGKNEPIYAKVANKK 899
Db 841 VKNVNGTLVNGSLGSGEATLAKKFSKIKELNEKFNKNNNNNGKNEPIYAKVANKK 900
Qy 900 KAGQAASLEERPIYAQVAKVNAKIDRLNQAASGLGVGQAGFPLKRHDKVDLDSKVGSRV 956
Db 901 KAGQAASLEERPIYAQVAKVNAKIDRLNQAASGLGVGQAGFPLKRHDKVDLDSKVGSRV 960
Qy 957 -----GLSRNOELAQKIDNLNQAASGLGVGQAGFPLKRHDKVDLDSKVGSRV 983
Db 957 -----GLSRNOELAQKIDNLNQAASGLGVGQAGFPLKRHDKVDLDSKVGSRV 983
Qy 961 ANHEPIYATIDDLGGPPFLSKRHDKVDLDSKVGSRVSPPIYATIDDLGGPPFLSKRHDKVD 1020
Db 961 ANHEPIYATIDDLGGPPFLSKRHDKVDLDSKVGSRVSPPIYATIDDLGGPPFLSKRHDKVD 1020
Qy 984 GNLEQFIDKLKSTKHNPMLNWSAKKVPASLSAKLDNATNSHTRNSLNKGAINEK 1043
Db 1021 DNLQDIDKLKSTKHNPMLNWSAKKVPASLSAKLDNATNSHTRNSLNKGAINEK 1080
Qy 1044 ATGMLTOKNPEWLKLVNDKIYAHNVGSPVLSYDKIGFNQNMKDYSDSFKESTKLNNAV 1103
Db 1081 ATGMLTOKNPEWLKLVNDKIYAHNVGSPVLSYDKIGFNQNMKDYSDSFKESTKLNNAV 1140
Qy 1104 KDTNSGFTQFLNAFSTASYICLARENAEHGKKNVTNKGFGOKS 1147
Db 1141 KDVKSQFTFLANAFST-GYCYLTEENAEHGIKNVTNKGFGOKS 1183

RESULT 9
ID 09F230 PRELIMINARY; PRT: 1163 AA.
AC 09F230;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CAGA.
OS CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
```



```

[1]
RN      SEQUENCE FROM N.A.
RC      STRAIN-J-187;
RA      Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA      Ando T.;
RT      "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT      Japanese and non-Japanese isolates.";
RL      J. Gastroenterol. 35:890-897(2000).
DR      EMBL: AB015406; BAB20916.1; -
SQ      SEQUENCE 1163 AA; 130090 MW; FC0AF8E3ED2845FC CRC64;

Query Match      80.88; Score 4741; DB 2; Length 1163;
Best Local Similarity 78.8%; Pred. No. 6.4e-191;
Matches 928; Conservative 94; Mismatches 111; Indels 44; Gaps 6;

Qy 1 MTNETIDQ----QPOTEAAFNPPQFINNLQVAFKVDNAVASYDDPQKPIVDKNDNRN 55
Db 1 MTNETIDQTTTPODTNQTDEVPQFINNLQVAFKVDNAVASYDDPQKPIVDKNDNRN 60
Qy 56 QAFEGISQLEEYSNKAIAKNTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQ 115
Db 61 QAFEEKISQLEMEYANKAIAKNTKKNQYFSDFINKSNLDINKNDLIAVDSSVESFRKFGDQ 120
Qy 116 RYRIFTSWVSHQNDPSKINTRSIRNFMENTIQQPILDDEKAEFLKSAKOSFAGIIGNQ 175
Db 121 RYQIFTSWVSLQKDPKINTQOIRNFMENTIQQPISDDKEAEFLRSKOSFAGIIGNQ 180
Qy 176 IRTDQKFMGVFDESLSKERQAEKNGEPTGGDWLDFISLFDKSSDVKAEINQEPVPH 235
Db 181 IRSDEKFMGVFDESLSKARQAEKNAEPAGGDWLDIFLSFVFNKKQSSDLKETLQOEPRD 240
Qy 236 VQPDIAATTTDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIH 295
Db 241 FEQNLAATTTDIOGLPPEARDLDERGNFEKFTLGDMEMLDVEGVADKDPNPKFNQLLIH 300
Qy 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGGPGARHDWATVGYKQDQGNVATINVMKN 355
Db 301 NNALSSVLMGSHSIEPEKISLLYGGNGGPGARHDWATVGYKQDQGNVATLINAHNN 360
Qy 356 GSGLVIAAGGEGKINPFSFYLYKEDQLNGSORALSQEEIQNKIDPMEFLAQNNAKLNLSE 415
Db 361 GSGLIAGNEDGINKPSFYLYKEDQLTGLKQALSQEEIQNKVDPMFLAQNNAKLNLSE 420
Qy 416 KEKEKFRTEIKDFOKSKAYLDALGNDRIFAIVSKKDKTKHSAITFEFGNGDLSYTLKDYGK 475
Db 421 KEKEKFOTEIENQKARKAYLDALGNDRIFAIVSKKDKTKHSAITFEFGNGELSYTLKDYGK 480
Qy 476 KADKALDREKNVTLOGSLKHGDNVFDYVSNPKYTNASKNPKNGVGTNGVSHLEVGNKY 535
Db 481 KQDKALDRETKTTLQGLKHKHGDGMFVFNYSNPKYTNASKNPKNGLGTNGVSHLEAFNSKV 540
Qy 536 AIFNLPLNLAITSFVRRLNLEOKLTITKGLSPOEANKLIKDFLSSNKELVGTILNFKAV 595
Db 541 AVFKPLNLAITNYIRROLEDKLAAGLSPOEASKLIKDFLNSKELLGKVSFNFKAV 600
Qy 596 ADAKNTGNYDEVKKAQKDLKSLRKREHLEKEVEKLEKSGNKNMEAKAQAANSQKDEI 655
Db 601 AEAKNTGNYDEVKKAQKDLKSLRKREHLEKEVEKLVVKKLESNDKNRMEAKAQAANSQKDKI 660
Qy 656 PALINKANDARAIAQAQNLKIGIKRELSDKLENVKNLKDQKSFDFPKNGKNKDFSKA 715
Db 661 FALINQASKEARAVAPDPSLKGIRSELSDKLENINKNLKDFGKSFDELKNGKNDPSKA 720
Qy 716 EETLKALKGSVKDGLGNPEWISKVENLNAALNEFKNKNKDKFSKVTOAKSDLENSVKDVI 775
Db 721 EETLKALKGSVKDGLGNPEWISKVENLNAALNDFKNGKNKDKFSKVTOAKSDLENSIKDVI 780
Qy 776 INQKVTDKVDNLNQAVSVAKATGDFSVEQALADLKNFSQEQLAQAQAQKNESLNARKKE 835
Db 781 INQKITDKVDNLNQAVSETKLTGDFSKEVEQALAEKLN-----SLDLGKNSD 827
Qy 836 IYQSVKNGVNGTLVGNGLSQAEATTLTKNSFSDIKKELNAKLGNFNNNNNGLKN--EPIY 893
```

```

Db 828 LQSVKNGVNGTLVGNGLSKTEATTLTKNFSDIRKELNEKLFNGSNNNNNGLKNNTPIY 887
Qy 894 AKVNKKKAGQAASLEPIYAQVAKKVNKIDRLNIOIAS-GLGVVG----- 937
Db 888 AQVNKKKTGQATSPEPIYAQVAKKMSAKIDRINKIASAGKGVGGFSGAGRSASPEPIYA 947
Qy 938 -----QAAGEPLKRRHDKVDLDSKVLGRNQELAQKIDNLNQAVSEAKAGFFGNGLEQTI 990
Db 948 TIDFEANQAGFPLRRSAGVNDLSKVLGRSQELTRRIGDLSQAVSEAKTGHFIDNLEQKI 1007
Qy 991 DKLDSTKHNPMNLWVESAKKVPASLSAKLONATNSHIRNSNTKNGAINKATGMLTQ 1050
Db 1008 DELKDSSTKKNALKLWESAKQVPTSLQAKLONATNSHTRINSNVQGTINEKATGMLTQ 1067
Qy 1051 KNPEWLKLVNDKIYAHNVGSLPYSDYKIGFNQKNMKDYSDSFKSTKLNNNAVKDNTSGF 1110
Db 1068 KNPEWLKLVNDKIYAHNVGSIHLSEYDKIGFNQKNMKDYSDSFKSTKLNNNVKDKISF 1127
Qy 1111 TQFLTNAFSTASYCLARENAEHGKKNVTNKGFFOKS 1147
Db 1128 VQFLTNTFTGCS-YSLMKANAEGHVKNTNKGFFOKS 1163

RESULT 10
Q9F229 PRELIMINARY; PRT; 1171 AA.
AC Q9F229;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J-198;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL: AB015407; BAB20917.1; -
SQ SEQUENCE 1171 AA; 130644 MW; E7FD06092C9E5DPE CRC64;

Query Match      80.4%; Score 4720; DB 2; Length 1171;
Best Local Similarity 78.3%; Pred. No. 4.9e-190;
Matches 929; Conservative 92; Mismatches 110; Indels 56; Gaps 7;
```

```

Qy 1 MTNETIDQ----QPOTEAAFNPPQFINNLQVAFKVDNAVASYDDPQKPIVDKNDNRNQA 57
Db 1 MTNETIDQTTTPODTNQTDEVPQFINNLQVAFKVDNAVASYDDPQKPIVDKNDNRNHA 58
Qy 58 FEGISQLEEYSNKAIAKNTKKNQYFSDFINKSNLDINKNDLIDVESSTKSFQKFGDQY 117
Db 59 FEKISQLEEYANKAIAKNTKKNQYFSDFINKSNLDINKNDLIAVDSSVESFRKFGDQY 118
Qy 118 RIFTSWVSHQNDPSKINTRSIRNFMENTIQQPILDDEKAEFLKSAKOSFAGIIGNOIR 177
Db 119 QIFTSWVSLQKDPKINTQOIRNFMENTIQQPISDDKEAEFLRSKOSFAGIIGNOIR 178
Qy 178 TDQKFMGVFDESLSKERQAEKNGEPTGGDWLDFISLFDKSSDVKAEINQEPVPHVQ 237
Db 179 SDEKFMGVFDESLSKARQAEKNAEPAGGDWLDIFLSFVFNKKQSSDLKETLQOEPRDPE 238
Qy 238 PDATTTTIDQGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNQLLIHNN 297
Db 239 QNLATTTTIDQGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADKDPNPKFNQLLIHNN 298
```



```
QY 298 ALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAIVGYKQQGNNVATIIIVHMKNGS 357
Db 299 ALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAIVGYKQQGNNVATIIIVHMKNGS 358
QY 358 GLVIAGGEGINNPFYLLKEDQGTGSRALSOEETQNKIDWFEFLAQNNAKLDNLSEKE 417
Db 359 GLUIAGNEDGINKNPFYLYKTDLTGKALSOEETQNKIDWFEFLAQNNAKLDNLSEKE 418
QY 418 KEKFTTEIKDFKOSKAYLDALGNDRIFAFVSKKDKHSAITIEFGNGDLSYTLKDYGKKA 477
Db 419 KEKFTTEIKDFKOSKAYLDALGNDRIFAFVSKKDKHSAITIEFGNGDLSYTLKDYGKKA 478
QY 478 DKALDREKNTVLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 537
Db 479 DKALDREKNTVLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 538
QY 538 FNLPLDNLNLAITSFVRNLEDEKLTGKLSPOEANKLIDFLSNKELVGKTLNFKAVAD 597
Db 539 FNLPLDNLNLAITSFVRNLEDEKLTGKLSPOEANKLIDFLSNKELVGKTLNFKAVAD 598
QY 598 AKNTGNYDEVKKAQKDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 657
Db 599 AKNTGNYDEVKKAQKDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 658
QY 658 LINKANRDRARATAYAOQLKIGIKRELSDKLENNVKNLKDPSKQVTOAKSDLENSVKDVIIN 717
Db 659 LINKANRDRARATAYAOQLKIGIKRELSDKLENNVKNLKDPSKQVTOAKSDLENSVKDVIIN 718
QY 718 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSVKDVIIN 777
Db 719 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSVKDVIIN 778
QY 778 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLKNFSQEALAOQAQKNEISLNAKKSIEY 837
Db 779 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLKNFSQEALAOQAQKNEISLNAKKSIEY 838
QY 838 QSVKNGVNTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNNNNNNNGLKNEPIYAKVN 897
Db 839 QSVKNGVNTLVNGLSQAEATTLKSNFSDIKKELNAKLGNNNNNNNNGLKNEPIYAKVN 898
QY 898 KKKAGAAASLEPIYAOVAKV-----NAKIDRLNOIAS-GLGVVG----- 937
Db 899 KKKAGAAASLEPIYAOVAKV-----NAKIDRLNOIAS-GLGVVG----- 938
QY 938 -----QAAGPPLKRDHVDLSKYGLSRNOELAQKIDNLNOAVSEAKA 980
Db 939 -----QAAGPPLKRDHVDLSKYGLSRNOELAQKIDNLNOAVSEAKA 981
QY 981 GFEGNLEQIDKLDKSTKINPNMLWVESAKVPASLSAKLDNATYNSHTRINSINIKGAI 1040
Db 982 GFEGNLEQIDKLDKSTKINPNMLWVESAKVPASLSAKLDNATYNSHTRINSINIKGAI 1041
QY 1041 NEKATGMLTQKPEWUKLVNDKIVAHNVGSPVLSYEDKIGFNOKNMKYSDSFKESTKLN 1100
Db 1042 NEKATGMLTQKPEWUKLVNDKIVAHNVGSPVLSYEDKIGFNOKNMKYSDSFKESTKLN 1101
QY 1101 NAVKDTNSGTFQTLTNAFTASYCLARENAEHGINKVNTKGGFKS 1147
Db 1102 NAVKDTNSGTFQTLTNAFTASYCLARENAEHGINKVNTKGGFKS 1148
QY 1147 NAVKDKISFVQTLTNTFTSGS-YSLMKANAEGHVKNKTNTKGGFKS 1171
Db 1148 NAVKDKISFVQTLTNTFTSGS-YSLMKANAEGHVKNKTNTKGGFKS 1172

RESULT 11
Q9F218 PRELIMINARY; PRT: 1170 AA.
AC Q9F218;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
```

```
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J-566;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RT Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB017922; BAB20928.1; -
SQ SEQUENCE 1170 AA; 130829 MW; 59A23543FBA86E03 CRC64;

Query Match 80.2%; Score 4703.5; DB 2; Length 1170;
Best Local Similarity 77.9%; Pred. No. 2.4e-189;
Matches 925; Conservative 99; Mismatches 106; Indels 57; Gaps 8;

QY 1 MTNETIDQ---QPOTEAAPNQPFINNQLQVAFKVDNAVASYDDPKPIVDKNDNRDNRQA 57
Db 1 MTNETIDQTTTPQDT--FVPRFINNQLQVAFKVDNAVASYDDPKPIVDKNDNRDNRQA 58
QY 58 FEGISOLREYYSNKAIKNPTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFKQFGDQRY 117
Db 59 FEGISOLREYYSNKAIKNPTKKNOYFSDFINKSNLDINKNDLIDVESSTKSFKQFGDQRY 118
QY 118 RIFTSWVSHQNDPSKINSTRINPMENIIQPPILDDKEAEFLKSAKOSFAGIIGNOIR 177
Db 119 QIFTSWVSLQKDPKSKINTQIRNFMENIIQPPISDDKEAEFLKSAKOSFAGIIGNOIR 178
QY 178 TDQKFMGVFDESLEKROEAEKNGEPTGGDWLDIFLSIFDKKQSSDVKEAINOEPVPHVQ 237
Db 179 SDEKFMGVFDESLEKROEAEKNAEPAGGDWLDIFLSFVFNKKQSSDLKETLNEPRPDE 238
QY 238 PDIAATTTDTQGLPPEARDLLDERGNEFKFTLGDMEMLDVEGVADIDPNYKFNQLLHNN 297
Db 239 QNLATTTTDTQGLPPEARDLLDERGNEFKFTLGDMEMLDVEGVADIDPNYKFNQLLHNN 298
QY 298 ALSVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAIVGYKQQGNNVATIIIVHMKNGS 357
Db 299 ALSVLMGSHNGIEPEKVSLLYGGNGPGARHDWNAIVGYKQQGNNVATIIIVHMKNGS 358
QY 358 GLVIAGGEGINNPFYLLKEDQGTGSRALSOEETQNKIDWFEFLAQNNAKLDNLSEKE 417
Db 359 GLVIAGGEGINNPFYLLKEDQGTGSRALSOEETQNKIDWFEFLAQNNAKLDNLSEKE 418
QY 418 KEKFTTEIKDFKOSKAYLDALGNDRIFAFVSKKDKHSAITIEFGNGDLSYTLKDYGKKA 477
Db 419 KEKFTTEIKDFKOSKAYLDALGNDRIFAFVSKKDKHSAITIEFGNGDLSYTLKDYGKKA 478
QY 478 DKALDREKNTVLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 537
Db 479 DKALDREKNTVLOGSLKHDGVNFVDSNFKYTNASKNPNKGVGTNGVSHLEVGFNKVAI 538
QY 538 FNLPLDNLNLAITSFVRNLEDEKLTGKLSPOEANKLIDFLSNKELVGKTLNFKAVAD 597
Db 539 FNLPLDNLNLAITSFVRNLEDEKLTGKLSPOEANKLIDFLSNKELVGKTLNFKAVAD 598
QY 598 AKNTGNYDEVKKAQKDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 657
Db 599 AKNTGNYDEVKKAQKDLKSLRKRREHLEKEVEKKLESKSGNKNKMEAKAQAQNSQKDEIFA 658
QY 658 LINKANRDRARATAYAOQLKIGIKRELSDKLENNVKNLKDPSKQVTOAKSDLENSVKDVIIN 717
Db 659 LINKANRDRARATAYAOQLKIGIKRELSDKLENNVKNLKDPSKQVTOAKSDLENSVKDVIIN 718
QY 718 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSVKDVIIN 777
Db 719 TLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNDKFSKVTOAKSDLENSVKDVIIN 778
QY 778 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLKNFSQEALAOQAQKNEISLNAKKSIEY 837
Db 779 QKVTDKVDNLNOAVSAKATGDFSRVEQALADLKNFSQEALAOQAQKNEISLNAKKSIEY 838
```


RP SEQUENCE FROM N.A.
RC STRAIN=J-149;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RT Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
Japanese and non-Japanese isolates."
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL: AB015405; BAB20915.1; -
SQ SEQUENCE 1176 AA; 131209 MW; DA8D473D2E2C47EE CRC64;

Query Match 79.9%; Score 4690.5; DB 2; Length 1176;
Best Local Similarity 77.6%; Pred. No. 8.4e-189;
Matches 928; Conservative 87; Mismatches 117; Indels 59; Gaps 8;

QY 1 MTNETIDQ-----QPTEAAFNPOQFINNLOVAFKVDNAVASYPDPQKPIVDKNDNRN 55
DB 1 MTNETIDQTPDTPNQTFDVPQFINNLOVAFKVDNAVASYPDPQKPIVDKNDNRN 60
QY 56 QAFEGISQREYSNKAIRNPTKKNQYFSDFINKSNDLINKMDLIDVESKTSFQKFGDQ 115
DB 61 QAFEGISQREYSNKAIRNPTKKNQYFSDFINKSNDLINKMDLIDVESKTSFQKFGDQ 120
QY 116 RYRFTSWSHQNDPSKINTSRNFMENIOPPIIDDKEKAEFLKSAKOSFAGIIGNQ 175
DB 116 RYRFTSWSHQNDPSKINTSRNFMENIOPPIIDDKEKAEFLKSAKOSFAGIIGNQ 180
QY 236 VQPDIAITTTDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 295
DB 241 FEQLATTTTDDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 300
QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 355
DB 301 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 360
QY 356 GSGVLVAGEGKGNFNFYLYKEDQTLGSGRALSQERIQKIDFMEFLAONNAKLDNLSE 415
DB 361 GSGVLVAGENGKGNFNFYLYKEDQTLGSGRALSQERIQKIDFMEFLAONNAKLDNLSE 420
QY 416 KEKEKTEKTEKDFQKDSKAYLDALGNDRIAFVSKKDTFKHSALITEFGNGDLSYTLKDYGK 475
DB 421 KEIRKPESETENFQKNPKAVLDALGNDHIVFVSKDKPKHLALVTEFGNGEVSYTLKDYGK 480
QY 476 KADKALDREKNVTLOGSLKHGDMVFDVYSNFKYTNASKNPKNGVGTNGVSHLEVGNKV 535
DB 481 KQKALDGETKTTLOGSLKYDGYVMFYNSNFKYTNASKNPKNGVGTNGVSHLEVGNKV 540
QY 536 AIFNLPDLNLAITSFVRRNLEDKLTTKGLSPQEAANKLIKDFLSSNKELVYKTLNFKAV 595
DB 541 AVFNLPDLNLAITNIRRDLEDKLWAKGLSPQEAANKLIKDFLSSNKELVYKTLNFKAV 600
QY 596 ADAKNTGNYDEVKKAQDLKSLRKEHLEKEVEKKLESKSGNKNMEAKAQAQNSQKDEI 655
DB 601 AEAKNTGNYDEVKKAQDLKSLRKEHLEKEVEKKLESKSGNKNMEAKAQAQNSQKDEI 660
QY 656 FALINKEANDRAIAYAQNKGIRKLSKLENVKNLKDFOKSFDEFKNGKNKDFSKA 715
DB 661 FALINTEASKEARVAFDPLKGIKRSLSKLENINKNLKDFGKSEDELKNGKNDFSKA 720
QY 716 EETLKALGSKVDLGINPEWISVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDI 775
DB 721 EETLKALGSKVDLGINPEWISVENLNAALNEFKNGKNKDFSKVTOAKSDLENSKVDI 780
QY 776 INOKVTDKVDNLNQAVSAKATGDFSRVEQALADLNFSEKQALAAQQAQKVESLNARKKSE 835
DB 781 INOKITDKVDNLNQAVSETKLTGDFSKVEQALAEKLSL-----SLDLGKNSD 827
QY 836 IYQSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNAKLFNFNNNNNGLKN--EPYI 893
DB 836 IYQSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNAKLFNFNNNNNGLKN--EPYI 893

DB 828 LQKSVKNGVNTLVGNSLQSAEATTTLSKNFSDIKKELNEKLFNFNNNNNGLKNSTPIY 887
QY 894 AKVKKKAGQAASLEPEPIYAQVAKV-----NAKIDRLNQIAS-GLGVVG- 937
DB 888 AKVKKKGTGVASPEPEPIYAQVAKVSAKIDQLNEATSAINRKIDRINKIASACKGVGG 947
QY 938 -----QAAGFPLKRDHKKVDDLSKVLGSRNQELAQKIDNLNOAVS 976
DB 948 SGAGRSASPEPIYATIDFDEANQAGFLRRSTGVNDLSKVLGSRQELTRIGLDNOAVS 1007
QY 977 EAKAGFGNLEOTIDKLKDSKTHNPMNLWVESAKKVPASLAKLDNATNSHIRINSIK 1036
DB 1008 EAKTGYDNLEQKIDELKSTKKNALLLVESAKQVPTGLQAKLDNATNSHIRINSVH 1067
QY 1037 NGAINKATGMLTKQKNPEWLKLVNDKIVAHNVGSPVLSSEYDKIGFNQKMKDYSDSKFS 1096
DB 1068 NGAINKATGMLTKQKNPEWLKLVNDKIVAHNVGSAHLSSEYDKIGFNQKMKDYSDSKFS 1127
QY 1097 TKLNAVKTNSGFTQFLTNAFSTASYCLARENAEHGKIKNVNTKGGFQKS 1147
DB 1128 TKLNAVKTNSGFTQFLTNAFSTASYCLARENAEHGKIKNVNTKGGFQKS 1176
RESULT 14
Q9F228 PRELIMINARY; PRT: 1176 AA.
AC Q9F228;
DT 01-MAR-2001 (TEmBLrel. 16, Created)
DT 01-MAR-2001 (TEmBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEmBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_TaxID=210;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J-207;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
Japanese and non-Japanese isolates."
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL: AB015408; BAB20918.1; -
SQ SEQUENCE 1176 AA; 131015 MW; 9842E72146F61B1B CRC64;

Query Match 79.9%; Score 4690.5; DB 2; Length 1176;
Best Local Similarity 77.6%; Pred. No. 8.4e-189;
Matches 924; Conservative 93; Mismatches 115; Indels 59; Gaps 8;

QY 1 MTNETIDQ-----QPTEAAFNPOQFINNLOVAFKVDNAVASYPDPQKPIVDKNDNRN 55
DB 1 MTNETIDQTPDTPNQTFDVPQFINNLOVAFKVDNAVASYPDPQKPIVDKNDNRN 60
QY 56 QAFEGISQREYSNKAIRNPTKKNQYFSDFINKSNDLINKMDLIDVESKTSFQKFGDQ 115
DB 61 QAFEGISQREYSNKAIRNPTKKNQYFSDFINKSNDLINKMDLIDVESKTSFQKFGDQ 120
QY 116 RYRFTSWSHQNDPSKINTSRNFMENIOPPIIDDKEKAEFLKSAKOSFAGIIGNQ 175
DB 121 RYQIFTSWVSLQKDPKINTQOIRNFMENIOPPIIDDKEKAEFLKSAKOSFAGIIGNQ 180
QY 176 IRTDOKFMGVFDESLEKROEAKENGEPTGGDWLIDFLSFIDKKOSSDVKEALNQEPVPH 235
DB 181 IRSDEKFMGVFDESLEKROEAKENGEPTGGDWLIDFLSFIDKKOSSDVKEALNQEPVPH 240
QY 236 VQPDIAITTTDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 295
DB 241 FEQLATTTTDDIOGLPPEARDLDERGNFSKFTLGDMEMLDVEGVADIDPNYKFNOLLIIH 300
QY 296 NNALSSVLMGSHNGIEPEKVSLLYGGNGPGARHDNATVGYKDOQGNVATIIINVHKN 355

```
Db 301 NNALSSVLMGSHSIEPEKVSLSLYGONGGPEARHDMWATVGYKQQGNNVATLINVHLNN 360
Qy 356 GSGLVIAAGGEGKGNPNPFIYKEDQTLGQSORALSQOEIQNKIDFMEFLAQNNAKLDLSE 415
Db 361 GSGLIAGNEDGKGNPFIYKEDQTLGKQASQOEIQNKIDFMEFLAQNNAKLDLSE 420
Qy 416 KEKEKFEITEIDFQKDSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 475
Db 421 KEKEKFEITEIDFQKDSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 480
Qy 476 KADKALDREKNVTLQGSLEKHDGVNFVSNFYKNTNASKNPNKGVGTNGVSHLEVGNKV 535
Db 481 KQKALDGEIKTTLQGSLEKHDGVNFVSNFYKNTNASKNPNKGVGTNGVSHLEVGNKV 540
Qy 536 AIFNLPLNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKA 595
Db 541 AVENLPLNLNLAITNYIRRDLEKDLWAKGLSPQEAANKLIKDFLSSNKELVGTILNFNKA 600
Qy 596 ADAKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKLESKSGNKNMEAKQAANSOKDEI 655
Db 601 AEAKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKLESKSGNKNMEAKQAANSOKDKI 660
Qy 656 FALINKEANDARAIAAQAQNLGKIKRELSDKLENNVKNLKDQKSFDEFKNGKNKDFSKA 715
Db 661 FALINQASKEARAAAFDPFKGIRSELSDKLENNVKNLKDQKSFDEFKNGKNKDFSKA 720
Qy 716 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVI 775
Db 721 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSIKDVI 780
Qy 776 INQKVTDKVDNLNOAVSVAKATGDFSVEQALADLKNFSKEQLAQAQAKNESLNARKKSE 835
Db 781 INQKITDKVDNLNOAVSETKLADGFSKEQLAELKSL-----SLDLGKNSD 827
Qy 836 IYOSVKNVNGTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNNNNNNGLKN--EPIY 893
Db 828 LQSVKNGVNGTLVGNGLSKEATTTLTKNFSIRKELNEKLFNGSNNNNNGLKNSAEPIY 887
Qy 894 AKYNNKAGQAASLEPIYAQVAKV-----NAKIDRLNQIAS--GLGVVG- 937
Db 888 AQYNNKKTGOVAPSEPIYAQVAKVSAKIDQLNEATSAINKRIDRLNK TASAGKGVGGF 947
Qy 938 -----QAAGPLKRHDKVDDLKSVGLSRNOELAQKIDNLNOAVS 976
Db 948 SGAGRSASPEPIYATIDFDEANQAGFPLRRSAGVNDLSKVGLSREOBLTRRIGLSQAVS 1007
Qy 977 EAKAGFEGNLEQTLKLDSTKINPNLWVESAKKVPASLSAKLDNYATNHSIRINSNIK 1036
Db 1008 EAKTGHFDNLEQKIDELKSTKKNALKWVESAKQVPTGLQAKLDNYATNHSIRINSNVQ 1067
Qy 1037 NGAINKEATGMLTKQNPWELKLVNDKIVAHNVGVSPLSEYDKIGFNKNMKDYSDSPKFS 1096
Db 1068 TGITNEKATGMLTKQNPWELKLVNDKIVAHNVGVSAPLSEYDKIGFNKNMKDYSDSPKFS 1127
Qy 1097 TKLNNVADPNSTGFTOFLTNAFSTASYICLARENAEHGKIVNNTKGGFKS 1147
Db 1128 TKLNNVADIKSSFVQFLTNTSTGFS-YSLMKANAHEGVKN-TTKSGFKS 1176

RESULT 15
Q9F225 PRELIMINARY; PRT; 1177 AA.
AC Q9F225;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CAGA.
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
```

```
RN SEQUENCE FROM N. A.
RP STRAIN=J-241;
RA Hoshino F.B., Katayama K., Watanabe K., Takahashi S., Uchimura H.,
RA Ando T.;
RT "Heterogeneity found in the cagA gene of Helicobacter pylori from
RL Japanese and non-Japanese isolates.";
RL J. Gastroenterol. 35:890-897(2000).
DR EMBL; AB015411; BAB20921.1; -.
SQ SEQUENCE 1177 AA; 131346 MW; E1BC418998F410B5 CRC64;

Query Match 79.9%; Score 4689; DB 2; Length 1177;
Best Local Similarity 77.6%; Pred. No. 9,7e-189;
Matches 924; Conservative 94; Mismatches 115; Indels 58; Gaps 7;

Qy 1 MTNETIDQ-----QPQTEAFAFPNPOQFINNQVAFKLYVDNNAVSYDPQCKPIVDKNDNRNR 55
Db 1 MTNETIDQTTTPDQTPNQOTFVPOQFINNQVAFIKVDSAVASFPDQCKPIVDKNDNRNR 60
Qy 56 QAPEGISQLEEEYSNKAINKPTKKNOYFSDPINKSNDLINKNDLINKNDLIDVESSTKSFQKFGDQ 115
Db 61 QAPEKISQLEEEYANKAIKNPAKKNQYFSDPINKSNDLINKNDLINKNDLIDVESSEKFKFGDQ 120
Qy 116 RYRIFTSWVSHQNDPSKINTSRISIRNFMENIIPQIPDLDDKEKAEFLKSAQSFAGIIGNQ 175
Db 121 RYLNFTSWVSLQKDPKINTQIQIONFENIIPQIPDSDDKEKAEFLRSKQSFAGIIGNQ 180
Qy 176 IRTDQKFMGVDFDESLEKQEAENKGEPTGGDWLDFISFIDFKQSSDVKEAINEQVYPH 235
Db 181 IRSDEKFMGVDFDESLEKQEAENKGEPTGGDWLDFISFVFNKKOSSDLKETLNOEPRPS 240
Qy 236 VQPDIAITTTDIOCLPPEARDLIDERNFNSKFTLGDMEMLDVEGVADIDPNYKFNOLLIH 295
Db 241 VEQNIATTTDIOCLPPEARDLIDERNFNFETLGDVEMLDVEGVDPKDPNYKFNOLLIH 300
Qy 296 NNALSSVLMGSHNGIEPEKVSLSLYGNGGPGARHDMWATVGYKQQGNNVATLINVHLNN 355
Db 301 NNALSSVLMGSHSIEPEKVSLSLYGNGGPGARHDMWATVGYKQQGNNVATLINVHLNN 360
Qy 356 GSGLVIAAGGEGKGNPNPFIYKEDQTLGQSORALSQOEIQNKIDFMEFLAQNNAKLDLSE 415
Db 361 GSGLIAGNEDGKGNPFIYKEDQTLGKQASQOEIQNKIDFMEFLAQNNAKLDLSE 420
Qy 416 KEKEKFEITEIDFQKDSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 475
Db 421 KEKEKFEITEIDFQKDSKAYLDALGNDRIAFVSKDKTHSALITEFGNGDLSYTLKDYGK 480
Qy 476 KADKALDREKNVTLQGSLEKHDGVNFVSNFYKNTNASKNPNKGVGTNGVSHLEVGNKV 535
Db 481 KQKALDGEIKTTLQGSLEKHDGVNFVSNFYKNTNASKNPNKGVGTNGVSHLEVGNKV 540
Qy 536 AIFNLPLNLNLAITSFVRRNLEDKLTGKLSPOEANKLIKDFLSSNKELVGTILNFNKA 595
Db 541 AVENLPLNLNLAITNYIRRDLEKDLWAKGLSPQEAANKLIKDFLSSNKELVGTILNFNKA 600
Qy 596 ADAKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKLESKSGNKNMEAKQAANSOKDEI 655
Db 601 AEAKNTGNYDEVKKAQKDLKSLRKRHELEKEVEKLESKSGNKNMEAKQAANSOKDKI 660
Qy 656 FALINKEANDARAIAAQAQNLGKIKRELSDKLENNVKNLKDQKSFDEFKNGKNKDFSKA 715
Db 661 FALINQASKEARAAAFDPFKGIRSELSDKLENNVKNLKDQKSFDEFKNGKNKDFSKA 720
Qy 716 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSVKDVI 775
Db 721 EETLKALKSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLENSIKDVI 780
Qy 776 INQKVTDKVDNLNOAVSVAKATGDFSVEQALADLKNFSKEQLAQAQAKNESLNARKKSE 835
Db 781 INQKITDKVDNLNOAVSETKLADGFSKEQLAELKSL-----SLDLGKNSD 827
Qy 836 IYOSVKNVNGTLVGNGLSQAEATTLSKNFSDIKKELNAKLGNNNNNNGLKN--EPIY 893
```

Search completed: February 13, 2002, 14:11:58
Job time: 657 sec

THIS PAGE BLANK (USPTO)